

AN

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
4 April 2002 (04.04.2002)

PCT

(10) International Publication Number
WO 02/26825 A2(51) International Patent Classification⁷: C07K 14/705

(21) International Application Number: PCT/US01/30661

(22) International Filing Date:
28 September 2001 (28.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/236,546	29 September 2000 (29.09.2000)	US
60/240,589	13 October 2000 (13.10.2000)	US
60/242,322	20 October 2000 (20.10.2000)	US
60/242,223	20 October 2000 (20.10.2000)	US
60/245,855	3 November 2000 (03.11.2000)	US
60/245,900	3 November 2000 (03.11.2000)	US
60/247,587	9 November 2000 (09.11.2000)	US
60/249,343	15 November 2000 (15.11.2000)	US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). GRAUL, Richard, C. [US/US]; 682-29th Avenue, San Francisco, CA 94121 (US). WALIA, Narinder, K. [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). HAFALIA, April, J., A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). THORNTON, Michael [US/US]; 9 Medway Road, Woodside, CA 94062 (US). KALLICK, Deborah, A. [US/US]; 900 Olive Street, Menlo Park, CA 94025 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422

(US). KHAN, Farrah, A. [IN/US]; 3617 Central Road #102, Glenview, IL 60025 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). ARVIZU, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). ROOPA, Reddy [IN/US]; 1233 W. McKinley Avenue, #3, Sunnyvale, CA 94086 (US). NGUYEN, Dannel, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). LEE, Ernestine, A. [US/US]; 624 Kains Street, Albany, CA 94706 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). ISON, Craig, H. [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). WALSH, Roderick, T. [IE/GB]; 8 Boundary Court, St. Lawrence Road, Canterbury, Kent CT1 3EZ (GB). POLICKY, Jennifer, L. [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: G-PROTEIN COUPLED RECEPTORS

(57) Abstract: The invention provides human G-protein coupled receptors (GCREC) and polynucleotides which identify and encode GCREC. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of GCREC.

WO 02/26825 A2

G-PROTEIN COUPLED RECEPTORS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of G-protein coupled receptors and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of G-protein coupled receptors. The present invention further relates to the use of specific G-protein coupled receptors to identify molecules that are involved in modulating taste or olfactory sensation.

BACKGROUND OF THE INVENTION

Signal transduction is the general process by which cells respond to extracellular signals. Signal transduction across the plasma membrane begins with the binding of a signal molecule, e.g., a hormone, neurotransmitter, or growth factor, to a cell membrane receptor. The receptor, thus activated, triggers an intracellular biochemical cascade that ends with the activation of an intracellular target molecule, such as a transcription factor. This process of signal transduction regulates all types of cell functions including cell proliferation, differentiation, and gene transcription. The G-protein coupled receptors (GPCRs), encoded by one of the largest families of genes yet identified, play a central role in the transduction of extracellular signals across the plasma membrane. GPCRs have a proven history of being successful therapeutic targets.

GPCRs are integral membrane proteins characterized by the presence of seven hydrophobic transmembrane domains which together form a bundle of antiparallel alpha (α) helices. GPCRs range in size from under 400 to over 1000 amino acids (Strosberg, A.D. (1991) *Eur. J. Biochem.* 196:1-10; Coughlin, S.R. (1994) *Curr. Opin. Cell Biol.* 6:191-197). The amino-terminus of a GPCR is extracellular, is of variable length, and is often glycosylated. The carboxy-terminus is cytoplasmic and generally phosphorylated. Extracellular loops alternate with intracellular loops and link the transmembrane domains. Cysteine disulfide bridges linking the second and third extracellular loops may interact with agonists and antagonists. The most conserved domains of GPCRs are the transmembrane domains and the first two cytoplasmic loops. The transmembrane domains account, in part, for structural and functional features of the receptor. In most cases, the bundle of α helices forms a ligand-binding pocket. The extracellular N-terminal segment, or one or more of the three extracellular loops, may also participate in ligand binding. Ligand binding activates the receptor by inducing a conformational change in intracellular portions of the receptor. In turn, the large, third intracellular loop of the activated receptor interacts with a heterotrimeric guanine nucleotide binding

(G) protein complex which mediates further intracellular signaling activities, including the activation of second messengers such as cyclic AMP (cAMP), phospholipase C, and inositol triphosphate, and the interaction of the activated GPCR with ion channel proteins. (See, e.g., Watson, S. and S. Arkinstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 2-6;

- 5 Bolander, F.F. (1994) Molecular Endocrinology, Academic Press, San Diego CA, pp. 162-176;
Baldwin, J.M. (1994) Curr. Opin. Cell Biol. 6:180-190.)

GPCRs include receptors for sensory signal mediators (e.g., light and olfactory stimulatory molecules); adenosine, γ -aminobutyric acid (GABA), hepatocyte growth factor, melanocortins, neuropeptide Y, opioid peptides, opsins, somatostatin, tachykinins, vasoactive intestinal polypeptide family, and vasopressin; biogenic amines (e.g., dopamine, epinephrine and norepinephrine, histamine, glutamate (metabotropic effect), acetylcholine (muscarinic effect), and serotonin); chemokines; lipid mediators of inflammation (e.g., prostaglandins and prostanoids, platelet activating factor, and leukotrienes); and peptide hormones (e.g., bombesin, bradykinin, calcitonin, C5a anaphylatoxin, endothelin, follicle-stimulating hormone (FSH), gonadotropic-releasing hormone (GnRH), neurokinin, thyrotropin-releasing hormone (TRH), and oxytocin). GPCRs which act as receptors for stimuli that have yet to be identified are known as orphan receptors.

The diversity of the GPCR family is further increased by alternative splicing. Many GPCR genes contain introns, and there are currently over 30 such receptors for which splice variants have been identified. The largest number of variations are at the protein C-terminus. N-terminal and cytoplasmic loop variants are also frequent, while variants in the extracellular loops or transmembrane domains are less common. Some receptors have more than one site at which variance can occur. The splice variants appear to be functionally distinct, based upon observed differences in distribution, signaling, coupling, regulation, and ligand binding profiles (Kilpatrick, G.J. et al. (1999) Trends Pharmacol. Sci. 20:294-301).

25 GPCRs can be divided into three major subfamilies: the rhodopsin-like, secretin-like, and metabotropic glutamate receptor subfamilies. Members of these GPCR subfamilies share similar functions and the characteristic seven transmembrane structure, but have divergent amino acid sequences. The largest family consists of the rhodopsin-like GPCRs, which transmit diverse extracellular signals including hormones, neurotransmitters, and light. Rhodopsin is a photosensitive GPCR found in animal retinas. In vertebrates, rhodopsin molecules are embedded in membranous stacks found in photoreceptor (rod) cells. Each rhodopsin molecule responds to a photon of light by triggering a decrease in cGMP levels which leads to the closure of plasma membrane sodium channels. In this manner, a visual signal is converted to a neural impulse. Other rhodopsin-like GPCRs are directly involved in responding to neurotransmitters. These GPCRs include the receptors for adrenaline (adrenergic receptors), acetylcholine (muscarinic receptors), adenosine, galanin, and

35

glutamate (N-methyl-D-aspartate/NMDA receptors). (Reviewed in Watson, S. and S. Arkinstall (1994) The G-Protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 7-9, 19-22, 32-35, 130-131, 214-216, 221-222; Habert-Ortoli, E. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9780-9783.)

5 The galanin receptors mediate the activity of the neuroendocrine peptide galanin, which inhibits secretion of insulin, acetylcholine, serotonin and noradrenaline, and stimulates prolactin and growth hormone release. Galanin receptors are involved in feeding disorders, pain, depression, and Alzheimer's disease (Kask, K. et al. (1997) Life Sci. 60:1523-1533). Other nervous system rhodopsin-like GPCRs include a growing family of receptors for lysophosphatidic acid and other
10 lysophospholipids, which appear to have roles in development and neuropathology (Chun, J. et al. (1999) Cell Biochem. Biophys. 30:213-242).

 The largest subfamily of GPCRs, the olfactory receptors, are also members of the rhodopsin-like GPCR family. These receptors function by transducing odorant signals. Numerous distinct olfactory receptors are required to distinguish different odors. Each olfactory sensory neuron
15 expresses only one type of olfactory receptor, and distinct spatial zones of neurons expressing distinct receptors are found in nasal passages. For example, the RA1c receptor, which was isolated from a rat brain library, has been shown to be limited in expression to very distinct regions of the brain and a defined zone of the olfactory epithelium (Raming, K. et al. (1998) Receptors Channels 6:141-151). However, the expression of olfactory-like receptors is not confined to olfactory tissues. For example,
20 three rat genes encoding olfactory-like receptors having typical GPCR characteristics showed expression patterns not only in taste and olfactory tissue, but also in male reproductive tissue (Thomas, M.B. et al. (1996) Gene 178:1-5).

 Members of the secretin-like GPCR subfamily have as their ligands peptide hormones such as secretin, calcitonin, glucagon, growth hormone-releasing hormone, parathyroid hormone, and
25 vasoactive intestinal peptide. For example, the secretin receptor responds to secretin, a peptide hormone that stimulates the secretion of enzymes and ions in the pancreas and small intestine (Watson, supra, pp. 278-283). Secretin receptors are about 450 amino acids in length and are found in the plasma membrane of gastrointestinal cells. Binding of secretin to its receptor stimulates the production of cAMP.

30 Examples of secretin-like GPCRs implicated in inflammation and the immune response include the EGF module-containing, mucin-like hormone receptor (Emr1) and CD97 receptor proteins. These GPCRs are members of the recently characterized EGF-TM7 receptors subfamily. These seven transmembrane hormone receptors exist as heterodimers in vivo and contain between three and seven potential calcium-binding EGF-like motifs. CD97 is predominantly expressed in leukocytes and is
35 markedly upregulated on activated B and T cells (McKnight, A.J. and S. Gordon (1998) J. Leukoc.

The third GPCR subfamily is the metabotropic glutamate receptor family. Glutamate is the major excitatory neurotransmitter in the central nervous system. The metabotropic glutamate receptors modulate the activity of intracellular effectors, and are involved in long-term potentiation (Watson, supra, p.130). The Ca^{2+} -sensing receptor, which senses changes in the extracellular concentration of calcium ions, has a large extracellular domain including clusters of acidic amino acids which may be involved in calcium binding. The metabotropic glutamate receptor family also includes pheromone receptors, the GABA_B receptors, and the taste receptors.

Other subfamilies of GPCRs include two groups of chemoreceptor genes found in the nematodes Caenorhabditis elegans and Caenorhabditis briggsae, which are distantly related to the mammalian olfactory receptor genes. The yeast pheromone receptors STE2 and STE3, involved in the response to mating factors on the cell membrane, have their own seven-transmembrane signature, as do the cAMP receptors from the slime mold Dictyostelium discoideum, which are thought to regulate the aggregation of individual cells and control the expression of numerous developmentally-regulated genes.

GPCR mutations, which may cause loss of function or constitutive activation, have been associated with numerous human diseases (Coughlin, supra). For instance, retinitis pigmentosa may arise from mutations in the rhodopsin gene. Furthermore, somatic activating mutations in the thyrotropin receptor have been reported to cause hyperfunctioning thyroid adenomas, suggesting that certain GPCRs susceptible to constitutive activation may behave as protooncogenes (Parma, J. et al. (1993) Nature 365:649-651). GPCR receptors for the following ligands also contain mutations associated with human disease: luteinizing hormone (precocious puberty); vasopressin V_2 (X-linked nephrogenic diabetes); glucagon (diabetes and hypertension); calcium (hyperparathyroidism, hypocalcuria, hypercalcemia); parathyroid hormone (short limbed dwarfism); β_3 -adrenoceptor (obesity, non-insulin-dependent diabetes mellitus); growth hormone releasing hormone (dwarfism); and adrenocorticotropin (glucocorticoid deficiency) (Wilson, S. et al. (1998) Br. J. Pharmacol. 125:1387-1392; Stadel, J.M. et al. (1997) Trends Pharmacol. Sci. 18:430-437). GPCRs are also involved in depression, schizophrenia, sleeplessness, hypertension, anxiety, stress, renal failure, and several cardiovascular disorders (Horn, F. and G. Vriend (1998) J. Mol. Med. 76:464-468).

In addition, within the past 20 years several hundred new drugs have been recognized that are directed towards activating or inhibiting GPCRs. The therapeutic targets of these drugs span a wide range of diseases and disorders, including cardiovascular, gastrointestinal, and central nervous system disorders as well as cancer, osteoporosis and endometriosis (Wilson, supra; Stadel, supra). For example, the dopamine agonist L-dopa is used to treat Parkinson's disease, while a dopamine antagonist is used to treat schizophrenia and the early stages of Huntington's disease. Agonists and

antagonists of adrenoceptors have been used for the treatment of asthma, high blood pressure, other cardiovascular disorders, and anxiety; muscarinic agonists are used in the treatment of glaucoma and tachycardia; serotonin 5HT_{1D} antagonists are used against migraine; and histamine H₁ antagonists are used against allergic and anaphylactic reactions, hay fever, itching, and motion sickness (Horn, supra).

Recent research suggests potential future therapeutic uses for GPCRs in the treatment of metabolic disorders including diabetes, obesity, and osteoporosis. For example, mutant V2 vasopressin receptors causing nephrogenic diabetes could be functionally rescued in vitro by co-expression of a C-terminal V2 receptor peptide spanning the region containing the mutations. This result suggests a possible novel strategy for disease treatment (Schöneberg, T. et al. (1996) EMBO J. 15:1283-1291). Mutations in melanocortin-4 receptor (MC4R) are implicated in human weight regulation and obesity. As with the vasopressin V2 receptor mutants, these MC4R mutants are defective in trafficking to the plasma membrane (Ho, G. and R.G. MacKenzie (1999) J. Biol. Chem. 274:35816-35822), and thus might be treated with a similar strategy. The type 1 receptor for parathyroid hormone (PTH) is a GPCR that mediates the PTH-dependent regulation of calcium homeostasis in the bloodstream. Study of PTH/receptor interactions may enable the development of novel PTH receptor ligands for the treatment of osteoporosis (Mannstadt, M. et al. (1999) Am. J. Physiol. 277:F665-F675).

The chemokine receptor group of GPCRs have potential therapeutic utility in inflammation and infectious disease. (For review, see Locati, M. and P.M. Murphy (1999) Annu. Rev. Med. 50:425-440.) Chemokines are small polypeptides that act as intracellular signals in the regulation of leukocyte trafficking, hematopoiesis, and angiogenesis. Targeted disruption of various chemokine receptors in mice indicates that these receptors play roles in pathologic inflammation and in autoimmune disorders such as multiple sclerosis. Chemokine receptors are also exploited by infectious agents, including herpesviruses and the human immunodeficiency virus (HIV-1) to facilitate infection. A truncated version of chemokine receptor CCR5, which acts as a coreceptor for infection of T-cells by HIV-1, results in resistance to AIDS, suggesting that CCR5 antagonists could be useful in preventing the development of AIDS.

The involvement of some GPCRs in taste and olfactory sensation has been reported. Complete or partial sequences of numerous human and other eukaryotic sensory receptors are currently known. (See, e.g., Pilpel, Y. and D. Lancet (1999) Protein Sci. 8:969-977; Mombaerts, P. (1999) Annu. Rev. Neurosci. 22:487-509. See also, e.g., patents EP 867508A2; US 5,874,243; WO 92/17585; WO 95/18140; WO 97/17444; and WO 99/67282.) It has been reported that the human genome contains approximately one thousand genes that encode a diverse repertoire of olfactory receptors (Rouquier, S. et al. (1998) Nat. Genet. 18:243-250; Trask, B.J. et al. (1998) Hum. Mol. Genet. 7:2007-2020).

The discovery of new G-protein coupled receptors, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of G-protein coupled receptors.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, G-protein coupled receptors, referred to collectively as "GCREC" and individually as "GCREC-1," "GCREC-2," "GCREC-3," "GCREC-4," "GCREC-5," "GCREC-6," "GCREC-7," "GCREC-8," "GCREC-9," "GCREC-10," "GCREC-11," "GCREC-12," "GCREC-13," "GCREC-14," "GCREC-15," and "GCREC-16." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-16.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-16. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:17-32.

The invention additionally provides G-protein coupled receptors that are involved in olfactory and/or taste sensation. The invention further provides polynucleotide sequences that encode said G-protein coupled receptors.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting

of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ

ID NO:17-32, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected

from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. The method comprises a) exposing a sample comprising the
5 polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

10 Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of
15 a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a
20 pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid
25 sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group
30 consisting of SEQ ID NO:1-16. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino
35 acid sequence selected from the group consisting of SEQ ID NO:1-16, b) a polypeptide comprising a

naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides methods of using G-protein coupled receptors of the invention involved in olfactory and/or taste sensation, biologically active fragments thereof (including those having receptor activity), and amino acid sequences having at least 90% sequence identity therewith, to identify compounds that agonize or antagonize the foregoing receptor polypeptides. These compounds are useful for modulating, blocking and/or mimicking specific tastes and/or odors.

The present invention also relates to the use of olfactory and/or taste receptors of the invention, biologically active fragments thereof (including those having receptor activity), and polypeptides having at least 90% sequence identity therewith, in combination with one or more other olfactory and/or taste receptor polypeptides, to identify a compound or plurality of compounds that modulate, mimic, and/or block a specific olfactory and/or taste sensation.

The invention also relates to cells that express an olfactory or taste receptor polypeptide of the invention, a biologically active fragment thereof (including those having receptor activity), or a polypeptide having at least 90% sequence identity therewith, and the use of such cells in cell-based screens to identify molecules that modulate, mimic, and/or block specific olfactory or taste sensations.

Still further, the invention relates to a cell that co-expresses at least one olfactory or taste G-protein coupled receptor polypeptide of the invention, and a G-protein, and optionally one or more other olfactory and/or taste G-protein coupled receptor polypeptides, and the use of such a cell in screens to identify molecules that modulate, mimic, and/or block specific olfactory and/or taste sensations.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b)

hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

20

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"GCREC" refers to the amino acid sequences of substantially purified GCREC obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of GCREC. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

An "allelic variant" is an alternative form of the gene encoding GCREC. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides.

Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding GCREC include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as GCREC or a polypeptide with at least one functional characteristic of GCREC. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding GCREC, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding GCREC. The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent GCREC. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of GCREC is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms “amino acid” and “amino acid sequence” refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where “amino acid sequence” is recited to refer to a sequence of a naturally occurring protein molecule, “amino acid sequence” and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

“Amplification” relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of GCREC. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

The term “antibody” refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind GCREC polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or

oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize
5 the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on
10 the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No.
15 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH₂), which may improve a desired property, e.g.,
20 resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia
25 virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on
30 substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified
35 sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having

modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic GCREC, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding GCREC or fragments of GCREC may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
5	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
10	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
15	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
20	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be

assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of GCREC or the polynucleotide encoding GCREC which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:17-32 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:17-32, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:17-32 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:17-32 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:17-32 and the region of SEQ ID NO:17-32 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-16 is encoded by a fragment of SEQ ID NO:17-32. A fragment of SEQ ID NO:1-16 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-16. For example, a fragment of SEQ ID NO:1-16 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-16. The precise length of a fragment of SEQ ID NO:1-16 and the region of SEQ ID NO:1-16 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a

standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example,

over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a
5 length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

10 The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of
15 substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap
20 penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version
25 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

30 *Expect: 10*

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length,
35 for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for

instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

5 "Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

 The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely
10 resembles a human antibody, and still retains its original binding ability.

 "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized
15 after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be
20 varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

 Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about
25 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY;
30 specifically see volume 2, chapter 9.

 High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking
35 reagents are used to block non-specific hybridization. Such blocking reagents include, for instance,

sheared and denatured salmon sperm DNA at about 100-200 $\mu\text{g/ml}$. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of GCREC which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of GCREC which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of GCREC. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of GCREC.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a

functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

5 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

10 "Post-translational modification" of an GCREC may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of GCREC.

 "Probe" refers to nucleic acid sequences encoding GCREC, their complements, or fragments
15 thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target
20 DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

 Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100,
25 or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

 Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold
30 Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge
35 MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence.

This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence.

Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions

(UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

“Reporter molecules” are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An “RNA equivalent,” in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term “sample” is used in its broadest sense. A sample suspected of containing GCREC, nucleic acids encoding GCREC, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms “specific binding” and “specifically binding” refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope “A,” the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term “substantially purified” refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A “substitution” refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

“Substrate” refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A “transcript image” refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

“Transformation” describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods

well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The

presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human G-protein coupled receptors (GCREC), the polynucleotides encoding GCREC, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS

program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

5 Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are G-protein coupled receptors. For example, SEQ ID NO:2 is 39% identical to rat seven transmembrane G-protein coupled receptor (GenBank ID g5525078) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $1.2e-89$, which indicates the probability of obtaining the observed
10 polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a secretin family 7-transmembrane receptor domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and BLAST analyses provide further corroborative evidence that SEQ ID NO:2 is a G-protein coupled receptor. In an alternative example, SEQ ID NO:4 is 32% identical to
15 human seven transmembrane-domain receptor (GenBank ID g2117161) as determined by BLAST. (See Table 2.) The BLAST probability score is $3.2e-35$. SEQ ID NO:4 also contains a latrophilin/CL-1-like GPS domain and a secretin family 7-transmembrane receptor domain as determined by searching the HMM-based PFAM database. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:4 is a 7-transmembrane
20 G-protein coupled receptor. In an alternative example, SEQ ID NO:5 is 34% identical to murine oxytocin receptor (GenBank ID g1902964) as determined by BLAST. (See Table 2.) The BLAST probability score is $7.1e-21$. SEQ ID NO:5 also contains a 7-transmembrane receptor domain as determined by searching the HMM-based PFAM database. (See Table 3.) Data from BLIMPS analyses provide further corroborative evidence that SEQ ID NO:5 is a G-protein coupled receptor.
25 In an alternative example, SEQ ID NO:6 is 23% identical to a human cysteinyl leukotriene receptor (GenBank ID g5359718) as determined by BLAST. (See Table 2.) The BLAST probability score is $2.5e-21$. Data from BLIMPS analyses provide further corroborative evidence that SEQ ID NO:6 is a G-protein coupled receptor. In an alternative example, SEQ ID NO:8 is 32% identical to an opsin from the Mexican tetra, a blind cave fish, (GenBank ID g440626) as determined by BLAST. (See
30 Table 2.) The BLAST probability score is $2.2e-22$. SEQ ID NO:8 also contains a rhodopsin family 7-transmembrane receptor domain as determined by searching the HMM-based PFAM database. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:8 is a rhodopsin family G-protein coupled receptor. In an alternative example, SEQ ID NO:9 is 44% identical to a putative human neurotransmitter receptor
35 (GenBank ID g2465432) as determined by BLAST. The BLAST probability score is $5.2e-69$. SEQ

ID NO:9 also contains a rhodopsin family 7-transmembrane receptor domain as determined by searching the HMM-based PFAM database. (See Table 3.) Data from BLIMPS and BLAST analyses provide further corroborative evidence that SEQ ID NO:9 is a rhodopsin family G-protein coupled receptor. In an alternative example, SEQ ID NO:11 is 82% identical to Marmota marmota olfactory receptor (GenBank ID g5901488) as determined by BLAST. (See Table 2.) The BLAST probability score is 1.5e-101. SEQ ID NO:11 also contains a rhodopsin family 7-transmembrane receptor domain as determined by searching the HMM-based PFAM database. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:11 is a G-protein coupled receptor. In an alternative example, SEQ ID NO:12 is 64% identical to Homo sapiens olfactory receptor (GenBank ID g2792018) as determined by BLAST. (See Table 2.) The BLAST probability score is 8.4e-99. SEQ ID NO:12 also contains a rhodopsin family 7-transmembrane receptor domain as determined by searching the HMM-based PFAM database. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:12 is a G-protein coupled receptor. In an alternative example, SEQ ID NO:15 is 57% identical to chicken olfactory receptor 4 (GenBank ID g1246534) as determined by BLAST. (See Table 2.) The BLAST probability score is 1.1e-91. SEQ ID NO:15 also contains a 7-transmembrane receptor (rhodopsin family) domain as determined by searching the HMM-based PFAM database. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:15 is an olfactory receptor. SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:13-14, and SEQ ID NO:16 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-16 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:17-32 or that distinguish between SEQ ID NO:17-32 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 2536292F6 is the identification number of an Incyte cDNA sequence, and BRAINOT18 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 72051732V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g3738039) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_XXXXXX_N₁_N₂YYYY_N₃_N₄ represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYY is the number of the prediction generated by the algorithm, and N_{1,2,3...}, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FLXXXXXX_gAAAAA_gBBBBB_1_N is the identification number of a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
--------	--

GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
5 INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

10 Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

15 The invention also encompasses GCREC variants. A preferred GCREC variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the GCREC amino acid sequence, and which contains at least one functional or structural characteristic of GCREC.

The invention also encompasses polynucleotides which encode GCREC. In a particular 20 embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:17-32, which encodes GCREC. The polynucleotide sequences of SEQ ID NO:17-32, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

25 The invention also encompasses a variant of a polynucleotide sequence encoding GCREC. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding GCREC. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:17-30 32 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ

ID NO:17-32. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of GCREC.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding GCREC, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring GCREC, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode GCREC and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring GCREC under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding GCREC or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding GCREC and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode GCREC and GCREC derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding GCREC or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:17-32 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE

amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA
5 sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

10 The nucleic acid sequences encoding GCREC may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.)
15 Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et
20 al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060).
Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo
25 Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National
Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about
30 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence
35 into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode GCREC may be cloned in recombinant DNA molecules that direct expression of GCREC, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express GCREC.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter GCREC-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of GCREC, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple

naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding GCREC may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser. 7:215-223*; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser. 7:225-232*.) Alternatively, GCREC itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of GCREC, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active GCREC, the nucleotide sequences encoding GCREC or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding GCREC. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding GCREC. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding GCREC and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding GCREC and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques,

and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

5 A variety of expression vector/host systems may be utilized to contain and express sequences encoding GCREC. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding GCREC. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding GCREC can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding GCREC into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of GCREC are needed, e.g. for the production of antibodies, vectors which direct high level expression of GCREC may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

35 Yeast expression systems may be used for production of GCREC. A number of vectors

containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra;
5 Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of GCREC. Transcription of sequences encoding GCREC may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.*
10 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill,
15 New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding GCREC may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain
20 infective virus which expresses GCREC in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of
25 DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of
30 GCREC in cell lines is preferred. For example, sequences encoding GCREC can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a
35 selective agent, and its presence allows growth and recovery of cells which successfully express the

introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *ap^r* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding GCREC is inserted within a marker gene sequence, transformed cells containing sequences encoding GCREC can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding GCREC under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding GCREC and that express GCREC may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of GCREC using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on GCREC is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See,

e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

5 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding GCREC include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding GCREC, or any fragments thereof, may be cloned into a vector
10 for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for
15 ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

 Host cells transformed with nucleotide sequences encoding GCREC may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence
20 and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode GCREC may be designed to contain signal sequences which direct secretion of GCREC through a prokaryotic or eukaryotic cell membrane.

 In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of
25 the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture
30 Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

 In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding GCREC may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric GCREC protein
35 containing a heterologous moiety that can be recognized by a commercially available antibody may

facilitate the screening of peptide libraries for inhibitors of GCREC activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the GCREC encoding sequence and the heterologous protein sequence, so that GCREC may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled GCREC may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

GCREC of the present invention or fragments thereof may be used to screen for compounds that specifically bind to GCREC. At least one and up to a plurality of test compounds may be screened for specific binding to GCREC. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of GCREC, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which GCREC binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express GCREC, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing GCREC or cell membrane fractions which contain GCREC are then contacted with a test compound and binding, stimulation, or inhibition of activity of either GCREC or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the

assay may comprise the steps of combining at least one test compound with GCREC, either in solution or affixed to a solid support, and detecting the binding of GCREC to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor.

5 Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

GCREC of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of GCREC. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for GCREC activity, wherein GCREC is combined with at least one test compound, and the activity of GCREC in
10 the presence of a test compound is compared with the activity of GCREC in the absence of the test compound. A change in the activity of GCREC in the presence of the test compound is indicative of a compound that modulates the activity of GCREC. Alternatively, a test compound is combined with an in vitro or cell-free system comprising GCREC under conditions suitable for GCREC activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of
15 GCREC may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding GCREC or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of
20 human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by
25 homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and
30 the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding GCREC may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate
35 into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al.

Polynucleotides encoding GCREC can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding GCREC is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress GCREC, e.g., by secreting GCREC in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

10 THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of GCREC and G-protein coupled receptors. In addition, the expression of GCREC is closely associated with brain tissue. Therefore, GCREC appears to play a role in cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections. In the treatment of disorders associated with increased GCREC expression or activity, it is desirable to decrease the expression or activity of GCREC. In the treatment of disorders associated with decreased GCREC expression or activity, it is desirable to increase the expression or activity of GCREC.

Therefore, in one embodiment, GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial

insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, α_1 -antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia,

autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and tongavirus.

15 In another embodiment, a vector capable of expressing GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified GCREC in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of GCREC may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those listed above.

25 In a further embodiment, an antagonist of GCREC may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GCREC. Examples of such disorders include, but are not limited to, those cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections described above. In one aspect, an antibody which specifically binds GCREC may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express GCREC.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding GCREC may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GCREC including, but not limited to, those described above.

35 In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary

sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of GCREC may be produced using methods which are generally known in the art. In particular, purified GCREC may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind GCREC. Antibodies to GCREC may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with GCREC or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to GCREC have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of GCREC amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to GCREC may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc.

Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce GCREC-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for GCREC may also be generated. For example, such fragments include, but are not limited to, $F(ab)_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the $F(ab)_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between GCREC and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering GCREC epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for GCREC. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of GCREC-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple GCREC epitopes, represents the average affinity, or avidity, of the antibodies for GCREC. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular GCREC epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the GCREC-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of GCREC, preferably in active form, from the

antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of GCREC-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding GCREC, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding GCREC. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding GCREC. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding GCREC may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency

(Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in GCREC expression or regulation causes disease, the expression of GCREC from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in GCREC are treated by constructing mammalian expression vectors encoding GCREC and introducing these vectors by mechanical means into GCREC-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of GCREC include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). GCREC may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous

gene encoding GCREC from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to GCREC expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding GCREC under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding GCREC to cells which have one or more genetic abnormalities with respect to the expression of GCREC. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Cséte, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are

described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both incorporated by reference herein.

5 In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding GCREC to target cells which have one or more genetic abnormalities with respect to the expression of GCREC. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing GCREC to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with
10 ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92
15 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned
20 herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to
25 deliver polynucleotides encoding GCREC to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA,
30 resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for GCREC into the alphavirus genome in place of the capsid-coding region results in the production of a large number of GCREC-coding RNAs and the synthesis of high levels of GCREC in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a
35 persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN)

indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of GCREC into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding GCREC.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding GCREC. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2'O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding GCREC. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased GCREC expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding GCREC may be therapeutically useful, and in the treatment of disorders associated with decreased GCREC expression or activity, a compound which specifically promotes expression of the polynucleotide encoding GCREC may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding GCREC is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding GCREC are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding GCREC. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to

a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of GCREC, antibodies to GCREC, and mimetics, agonists, antagonists, or inhibitors of GCREC.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without

needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

5 Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising GCREC or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, GCREC or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to
10 transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and
15 route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example GCREC or fragments thereof, antibodies of GCREC, and agonists, antagonists or inhibitors of GCREC, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be
20 determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are
25 used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the
30 subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or
35 biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind GCREC may be used for the diagnosis of disorders characterized by expression of GCREC, or in assays to monitor patients being treated with GCREC or agonists, antagonists, or inhibitors of GCREC. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for GCREC include methods which utilize the antibody and a label to detect GCREC in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring GCREC, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of GCREC expression. Normal or standard values for GCREC expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to GCREC under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of GCREC expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding GCREC may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of GCREC may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of GCREC, and to monitor regulation of GCREC levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding GCREC or closely related molecules may be used to identify nucleic acid sequences which encode GCREC. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the

probe identifies only naturally occurring sequences encoding GCREC, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the GCREC encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:17-32 or from genomic sequences including promoters, enhancers, and introns of the GCREC gene.

Means for producing specific hybridization probes for DNAs encoding GCREC include the cloning of polynucleotide sequences encoding GCREC or GCREC derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding GCREC may be used for the diagnosis of disorders associated with expression of GCREC. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic

paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cardiovascular disorder such as

5 arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid

10 aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal

15 carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease,

20 Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis,

25 peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia,

30 autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome,

35 multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis,

osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and tongavirus. The polynucleotide sequences encoding GCREC may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered GCREC expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding GCREC may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding GCREC may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding GCREC in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of GCREC, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding GCREC, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding GCREC may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding GCREC, or a fragment of a polynucleotide complementary to the polynucleotide encoding GCREC, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of GCREC include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C.

et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the ligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

5 In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene
10 function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and
15 display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, GCREC, fragments of GCREC, or antibodies specific for GCREC may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

20 A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No.
25 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The
30 resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention
35 may also be used in conjunction with in vitro model systems and preclinical evaluation of

pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) *Mol. Carcinog.* 24:153-159; Steiner, S. and N.L. Anderson (2000)

5 Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested
10 compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for
15 example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample
20 containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are
25 indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are
30 analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl
35 sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins

are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for GCREC to quantify the levels of GCREC expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoz, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological

sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding GCREC may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding GCREC on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse,

may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any
5 sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, GCREC, its catalytic or immunogenic fragments, or
10 oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between GCREC and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds
15 having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with GCREC, or fragments thereof, and washed. Bound GCREC is then detected by methods well known in the art. Purified GCREC can also be coated directly onto plates for use in the aforementioned drug screening techniques.
20 Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding GCREC specifically compete with a test compound for binding GCREC. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more
25 antigenic determinants with GCREC.

In additional embodiments, the nucleotide sequences which encode GCREC may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

30 Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below,
35 including U.S. Ser. No. 60/236,546, U.S. Ser. No. 60/240,589, U.S. Ser. No. 60/242,223, U.S. Ser.

EXAMPLES

5 I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic
10 solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was
15 isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA
20 libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the
25 appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPO1 plasmid (Life Technologies), PCDNA2.1 plasmid
30 (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies.

35 II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, 5 QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal 10 cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

15 Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared 20 using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI 25 protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing 30 vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family 35 databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary

structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.)

The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER.

The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences.

Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or

- 5 Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may
- 10 begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering,
- 15 South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

- Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of
- 20 Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a
- 25 match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

- The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:17-32. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and
- 30 amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

- Putative G-protein coupled receptors were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a
- 35 variety of organisms (See Burge, C. and S. Karlin (1997) *J. Mol. Biol.* 268:78-94, and Burge, C. and

S. Karlin (1998) *Curr. Opin. Struct. Biol.* 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode G-protein coupled receptors, the encoded polypeptides were analyzed by querying against PFAM models for G-protein coupled receptors. Potential G-protein coupled receptors were also identified by homology to Incyte cDNA sequences that had been annotated as G-protein coupled receptors. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpr public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent

type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpr public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

5 "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST
10 analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous
15 genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of GCREC Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:17-32 were compared with
20 sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:17-32 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for
25 Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between
30 chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other
35 resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site

(<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding GCREC are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia,

male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding GCRC. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

10 VIII. Extension of GCRC Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the

concentration of DNA. A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviII cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:17-32 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10⁷ counts per minute of the labeled probe is used in a typical membrane-based

hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra.*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Skena (1999), *supra.*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Skena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first

strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μ l 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and

Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterpr of chamber having a cavity just slightly larger than a microsc pe slide. The chamber is kept at 100% humidity internally by the addition of 140
5 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an
10 Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a
15 resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The
20 emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on
25 the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and
30 adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high
35 signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and

measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the GCREC-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring GCREC. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of GCREC. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the GCREC-encoding transcript.

XII. Expression of GCREC

Expression and purification of GCREC is achieved using bacterial or virus-based expression systems. For expression of GCREC in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express GCREC upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of GCREC in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding GCREC by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, GCREC is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step,

affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from GCREC at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified GCREC obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

XIII. Functional Assays

GCREC function is assessed by expressing the sequences encoding GCREC at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of GCREC on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding GCREC and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions

of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding GCREC and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of GCREC Specific Antibodies

GCREC substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the GCREC amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-GCREC activity by, for example, binding the peptide or GCREC to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring GCREC Using Specific Antibodies

Naturally occurring or recombinant GCREC is substantially purified by immunoaffinity chromatography using antibodies specific for GCREC. An immunoaffinity column is constructed by covalently coupling anti-GCREC antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing GCREC are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of GCREC (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/GCREC binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and GCREC is collected.

XVI. Identification of Molecules Which Interact with GCREC

Molecules which interact with GCREC may include agonists and antagonists, as well as

molecules involved in signal transduction, such as G proteins. GCREC, or a fragment thereof, is labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) A fragment of GCREC includes, for example, a fragment comprising one or more of the three extracellular loops, the extracellular N-terminal region, or the third intracellular loop.

- 5 Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled GCREC, washed, and any wells with labeled GCREC complex are assayed. Data obtained using different concentrations of GCREC are used to calculate values for the number, affinity, and association of GCREC with the candidate ligand molecules.

Alternatively, molecules interacting with GCREC are analyzed using the yeast two-hybrid
10 system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech). GCREC may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S.
15 Patent No. 6,057,101).

Potential GCREC agonists or antagonists may be tested for activation or inhibition of GCREC receptor activity using the assays described in sections XVII and XVIII. Candidate molecules may be selected from known GPCR agonists or antagonists, peptide libraries, or combinatorial chemical libraries.

- 20 Methods for detecting interactions of GCREC with intracellular signal transduction molecules such as G proteins are based on the premise that internal segments or cytoplasmic domains from an orphan G protein-coupled seven transmembrane receptor may be exchanged with the analogous domains of a known G protein-coupled seven transmembrane receptor and used to identify the G-proteins and downstream signaling pathways activated by the orphan receptor domains (Kobilka, B.K.
25 et al. (1988) *Science* 240:1310-1316). In an analogous fashion, domains of the orphan receptor may be cloned as a portion of a fusion protein and used in binding assays to demonstrate interactions with specific G proteins. Studies have shown that the third intracellular loop of G protein-coupled seven transmembrane receptors is important for G protein interaction and signal transduction (Conklin, B.R. et al. (1993) *Cell* 73:631-641). For example, the DNA fragment corresponding to the third intracellular
30 loop of GCREC may be amplified by the polymerase chain reaction (PCR) and subcloned into a fusion vector such as pGEX (Pharmacia Biotech). The construct is transformed into an appropriate bacterial host, induced, and the fusion protein is purified from the cell lysate by glutathione-Sepharose 4B (Pharmacia Biotech) affinity chromatography.

- For in vitro binding assays, cell extracts containing G proteins are prepared by extraction with
35 50 mM Tris, pH 7.8, 1 mM EGTA, 5 mM MgCl₂, 20 mM CHAPS, 20% glycerol, 10 μg of both

aprotinin and leupeptin, and 20 μ l of 50 mM phenylmethylsulfonyl fluoride. The lysate is incubated on ice for 45 min with constant stirring, centrifuged at 23,000 g for 15 min at 4°C, and the supernatant is collected. 750 μ g of cell extract is incubated with glutathione S-transferase (GST) fusion protein beads for 2 h at 4°C. The GST beads are washed five times with phosphate-buffered saline. Bound
5 G protein subunits are detected by [32 P]ADP-ribosylation with pertussis or cholera toxins. The reactions are terminated by the addition of SDS sample buffer (4.6% (w/v) SDS, 10% (v/v) β -mercaptoethanol, 20% (w/v) glycerol, 95.2 mM Tris-HCl, pH 6.8, 0.01% (w/v) bromphenol blue). The [32 P]ADP-labeled proteins are separated on 10% SDS-PAGE gels, and autoradiographed. The separated proteins in these gels are transferred to nitrocellulose paper, blocked with blotto (5% nonfat
10 dried milk, 50 mM Tris-HCl (pH 8.0), 2 mM CaCl_2 , 80 mM NaCl, 0.02% NaN_3 , and 0.2% Nonidet P-40) for 1 hour at room temperature, followed by incubation for 1.5 hours with $\text{G}\alpha$ subtype selective antibodies (1:500; Calbiochem-Novabiochem). After three washes, blots are incubated with horseradish peroxidase (HRP)-conjugated goat anti-rabbit immunoglobulin (1:2000, Cappel, Westchester PA) and visualized by the chemiluminescence-based ECL method (Amersham Corp.).

15 XVII. Demonstration of GCREC Activity

An assay for GCREC activity measures the expression of GCREC on the cell surface. cDNA encoding GCREC is transfected into an appropriate mammalian cell line. Cell surface proteins are labeled with biotin as described (de la Fuente, M.A. et al. (1997) *Blood* 90:2398-2405). Immunoprecipitations are performed using GCREC-specific antibodies, and immunoprecipitated
20 samples are analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of GCREC expressed on the cell surface.

In the alternative, an assay for GCREC activity is based on a prototypical assay for ligand/receptor-mediated modulation of cell proliferation. This assay measures the rate of DNA
25 synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding GCREC is added to quiescent 3T3 cultured cells using transfection methods well known in the art. The transiently transfected cells are then incubated in the presence of [^3H]thymidine, a radioactive DNA precursor molecule. Varying amounts of GCREC ligand are then added to the cultured cells. Incorporation of [^3H]thymidine into acid-precipitable DNA is measured over an appropriate time interval using a
30 radioisotope counter, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold GCREC ligand concentration range is indicative of receptor activity. One unit of activity per milliliter is defined as the concentration of GCREC producing a 50% response level, where 100% represents maximal incorporation of [^3H]thymidine into acid-precipitable DNA (McKay, I. and I. Leigh, eds. (1993)
35 Growth Factors: A Practical Approach, Oxford University Press, New York NY, p. 73.)

In a further alternative, the assay for GCREC activity is based upon the ability of GPCR family proteins to modulate G protein-activated second messenger signal transduction pathways (e.g., cAMP; Gaudin, P. et al. (1998) J. Biol. Chem. 273:4990-4996). A plasmid encoding full length GCREC is transfected into a mammalian cell line (e.g., Chinese hamster ovary (CHO) or human embryonic kidney (HEK-293) cell lines) using methods well-known in the art. Transfected cells are grown in 12-well trays in culture medium for 48 hours, then the culture medium is discarded, and the attached cells are gently washed with PBS. The cells are then incubated in culture medium with or without ligand for 30 minutes, then the medium is removed and cells lysed by treatment with 1 M perchloric acid. The cAMP levels in the lysate are measured by radioimmunoassay using methods well-known in the art. Changes in the levels of cAMP in the lysate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

To measure changes in inositol phosphate levels, the cells are grown in 24-well plates containing 1×10^5 cells/well and incubated with inositol-free media and [^3H]myoinositol, 2 μCi /well, for 48 hr. The culture medium is removed, and the cells washed with buffer containing 10 mM LiCl followed by addition of ligand. The reaction is stopped by addition of perchloric acid. Inositol phosphates are extracted and separated on Dowex AG1-X8 (Bio-Rad) anion exchange resin, and the total labeled inositol phosphates counted by liquid scintillation. Changes in the levels of labeled inositol phosphate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

XVIII. Identification of GCREC Ligands

GCREC is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293 which have a good history of GPCR expression and which contain a wide range of G-proteins allowing for functional coupling of the expressed GCREC to downstream effectors. The transformed cells are assayed for activation of the expressed receptors in the presence of candidate ligands. Activity is measured by changes in intracellular second messengers, such as cyclic AMP or Ca^{2+} . These may be measured directly using standard methods well known in the art, or by the use of reporter gene assays in which a luminescent protein (e.g. firefly luciferase or green fluorescent protein) is under the transcriptional control of a promoter responsive to the stimulation of protein kinase C by the activated receptor (Milligan, G. et al. (1996) Trends Pharmacol. Sci. 17:235-237). Assay technologies are available for both of these second messenger systems to allow high throughput readout in multi-well plate format, such as the adenylyl cyclase activation FlashPlate Assay (NEN Life Sciences Products), or fluorescent Ca^{2+} indicators such as Fluo-4 AM (Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In cases where the physiologically relevant second messenger pathway is not

known, GCREC may be coexpressed with the G-proteins $G_{\alpha 15/16}$ which have been demonstrated to couple to a wide range of G-proteins (Offermanns, S. and M.I. Simon (1995) J. Biol. Chem. 270:15175-15180), in order to funnel the signal transduction of the GCREC through a pathway involving phospholipase C and Ca^{2+} mobilization. Alternatively, GCREC may be expressed in
5 engineered yeast systems which lack endogenous GPCRs, thus providing the advantage of a null background for GCREC activation screening. These yeast systems substitute a human GPCR and G_{α} protein for the corresponding components of the endogenous yeast pheromone receptor pathway. Downstream signaling pathways are also modified so that the normal yeast response to the signal is converted to positive growth on selective media or to reporter gene expression (Broach, J.R. and J.
10 Thorner (1996) Nature 384 (supp.):14-16). The receptors are screened against putative ligands including known GPCR ligands and other naturally occurring bioactive molecules. Biological extracts from tissues, biological fluids and cell supernatants are also screened.

Various modifications and variations of the described methods and systems of the invention
15 will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the
20 following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
2536292	1	2536292CD1	17	2536292CB1
7477708	2	7477708CD1	18	7477708CB1
7474823	3	7474823CD1	19	7474823CB1
644692	4	644692CD1	20	644692CB1
3837054	5	3837054CD1	21	3837054CB1
6157025	6	6157025CD1	22	6157025CB1
55012817	7	55012817CD1	23	55012817CB1
7475061	8	7475061CD1	24	7475061CB1
7477374	9	7477374CD1	25	7477374CB1
7479890	10	7479890CD1	26	7479890CB1
7482825	11	7482825CD1	27	7482825CB1
7483087	12	7483087CD1	28	7483087CB1
7483134	13	7483134CD1	29	7483134CB1
7478550	14	7478550CD1	30	7478550CB1
7483142	15	7483142CD1	31	7483142CB1
7483151	16	7483151CD1	32	7483151CB1

Table 2

Polyptide SEQ ID NO:	Incyte Polyptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	2536292CD1	g1039470	8.1E-07	[Rattus norvegicus] pheromone receptor VN1 (Dulac, C. and R. Axel (1996) Cell 83:195-206)
2	7477708CD1	g5525078	1.2E-89	[Rattus norvegicus] seven transmembrane receptor (Abe, J. et al. (1999) J. Biol. Chem. 274:19957-19964)
3	7474823CD1	g10241847	6.0E-79	[Homo sapiens] histamine H4 receptor (Oda, T. et al. (2000) J. Biol. Chem. 275:36781-36786)
4	644692CD1	g2117161	3.2E-35	[Homo sapiens] seven transmembrane-domain receptor (Osterhoff, C. et al. (1997) DNA Cell Biol. 16:379-389)
5	3837054CD1	g1902964	7.1E-21	[Mus sp.] oxytocin receptor (Kubota, Y. (1996) Mol. Cell. Endocrinol. 124:25-32)
6	6157025CD1	g5359718	2.5E-21	[Homo sapiens] cysteinyl leukotriene receptor (Sarau, H.M. et al. (1999) Mol. Pharmacol. 56:657-663)
7	55012817CD1	g6006811	8.6E-60	[Mus musculus] serpentine receptor
8	7475061CD1	g440626	2.2E-22	[Astyanax mexicanus] opsin
9	7477374CD1	g14600082	0.0	[Homo sapiens] trace amine receptor 3 (Borowsky, B. et al. (2001) Proc. Natl. Acad. Sci. USA 98:8966-8971)
10	7479890CD1	g10441732	0.0	[Homo sapiens] leucine-rich repeat-containing G protein-coupled receptor 6 (Hsu, S.Y. et al. (2000) Mol. Endocrinol. 14:1257-1271)
11	7482825CD1	g4680254	1.3E-68	[Mus musculus] odorant receptor S1 (Hirono, M.B. et al. (1999) Cell 96:713-723)
12	7483087CD1	g5901488 g2792018	1.5E-101 8.4E-99	[Marmota marmota] olfactory receptor [Homo sapiens] olfactory receptor (Vanderhaeghen, P. et al. (1997) Biochem. Biophys. Res. Commun. 237:283-287)
13	7483134CD1	g6178008	8.3E-92	[Mus musculus] odorant receptor MOR18 (Tsuboi, A. et al. (1999) J. Neurosci. 19:8409-8418)
14	7478550CD1	g3983394	1.2E-53	[Mus musculus] olfactory receptor F7 (Krautwurst, D. et al. (1998) Cell 95:917-926)
15	7483142CD1	g1246534	1.1E-91	[Gallus gallus] olfactory receptor 4 (Leibovici, M. et al. (1996) Dev. Biol. 175:118-131)
16	7483151CD1	g1246532	1.9E-78	[Gallus gallus] olfactory receptor 3 (Leibovici, M. et al. (1996) Dev. Biol. 175:118-131)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	2536292CD1	217	S183 S53 T172 T206 T45 Y128	N117 N181	Transmembrane domain: L63-L88	HMMER
2	7477708CD1	578	S179 S196 S224 S248 S255 S570 T185 T190 T346 T46 T499 T539 T97 S4 T16 T51	N122 N130 N175 N183 N228 N90	Transmembrane domains: Y318-V341, L400-I419, I479-V497 7 transmembrane receptor (secretin family): S313-V560 Latrophilin/CL-1-like GPS domain: R260-L311 Secretin-like GPCR super family: I518-A538, S376-F397, Y318-W342, T9-V33, A390-I413 G-protein coupled receptor PD000752: S307-Q559	HMMER HMMER-PFAM HMMER-PFAM
3	7474823CD1	441	S174 S209 S215 S232 S314 S322 S359 S415 T129 T320 T363 T428 T433	N324 N393 N6	EMR1 hormone receptor: DM05221 I37225 347-738: C264-Q559 DM05221 P48960 347-738: C265-Q559 DM05221 A57172 465-886: V231-F558 Signal peptide: M1-A51 Transmembrane domain: Y226-N247 7 transmembrane receptor (rhodopsin family): F131-W255 G-protein coupled receptor motif: A153-V169 G-protein coupled receptors signature: D147-V194 G-protein coupled receptor BL00237: W133-A172, F236-N247, T177-M203 Rhodopsin-like GPCR superfamily: PR00237: K184-L205, L228-Y251, V182-V206, W225-Y251, K184-E208, D147-V169	SPScan HMMER HMMER-PFAM MOTIFS ProfileScan BLIMPS-BLOCKS BLIMPS-PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3					G-protein coupled receptor: DM00013 P30546 22-383: V119-H259 DM00013 P22270 103-599: S118-R257 DM00013 S58868 45-481: S118-R257 DM00013 P31390 21-315: I120-G281	BLAST-DOMO
4	644692CD1	797	S249 S3 S344 S349 S350 S368 S389 S466 S492 S499 S647 S660 S665 S666 S696 S746 S781 S788 T192 T225 T233 T294 T338 T423 T459 T47 T558 T661 T721 T768 T786	N159 N178 N191 N247 N261 N312 N316 N387 N413 N657 N709 N82	Egf: C63-C74 Signal peptide: M1-G26 Transmembrane domain: N8-W30, V470-I489, I567-Q592, V632-W652, S672-L689 7 transmembrane receptor (Secretin family) 7tm_2: P431-V726 Iatrophilin/CL-1-like GPS domain GPS: Y379-D427 G-protein coupled receptor BL00649: S389-T416, G441-V486, C526-L551, G574-Q598, W626-N655, S666-A687, T710-L735 C.elegans integral membrane protein Srb signature PR00699E: V467-F487 HORMONE; EMR1; LEUCOCYTE; ANTIGEN: DM05221 A57172 465-886: C526-Y720 DM05221 I37225 347-738: Y603-T751 GPROTEIN COUPLED TRANSMEMBRANE RECEPTOR PD000752: N511-E730 G-PROTEIN COUPLED RECEPTORS: DM00013 P47901 28-353: R43-V182, W260-A357, W227-L256 G-protein coupled receptor BL00237: A115-P154, R288-A314, N344-C360 Transmembrane domain: R45-C65 7-transmembrane receptor (rhodopsin family, 7tm_1): G59-F181, H228-Y352 Signal cleavage: M1-A58	MOTIFS HMME HMME HMME-PFAM HMME-PFAM
5	3837054CD1	434	S41 S131 S293	N15 N27 N60		BLIMPS-PRINTS BLAST-DOMO BLAST-PRODROM BLAST-DOMO BLIMPS-BLOCKS HMME HMME-PFAM SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	6157025CD1	339	T6 S10 Y175 S217 S276 S332	N5 N9 N308	G-PROTEIN COUPLED RECEPTORS: DM00013 P34993 36-327: L20-F285 DM00013 P30872 52-338: L20-F285 G-protein coupled receptor BL00237: H218-F244, N9-F25, W81-C120 Transmembrane domain: I21-M45, V43-S63, M99-I116, M139-Y155, I185-V208, N227-L249 Signal peptide: M1-G20 Signal peptide: M1-E22 Transmembrane domains: G499-I523, L477-F495, H436-F458, Y378-L405, A274-L293 7 transmembrane receptor (Secretin family): V265-S528 G-protein coupled receptor BL00649: M280-V325, C338-L363, G386-D410, Y429-F458, N472-A493 Secretin-like GPCR superfamily PR00249: R270-R294, A340-L363, F379-G404, L492-I517, K475-F495, V503-L524 EMR1 leucocyte antigen: DM05221 I37225 347-738: P207-I517 DM05221 P48960 347-738: P207-I517 DM05221 A57172 465-886: L177-I517 G-protein coupled receptor PD000752: I271-L524 Signal peptide: M1-I50 Transmembrane domains: F34-F52, M90-F112	BLAST-DOMO BLIMPS-BLOCKS HMMER SPScan HMMER HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DOMO BLAST-PRODROM SPScan HMMER
7	55012817CD1	549	S19 S57 T103 T153 T224 T3 T407 T415 T537 T84	N144 N210 N413 N98		
8	7475061CD1	188	S163 S179 S64 T159 T174			

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8					7 transmembrane receptor (rhodopsin family): I38-Y141 Visual pigments (opsins) retinal binding site: A130-F148 Visual pigments (opsins) retinal binding site: A111-G166 G-protein coupled receptor BL00237: P13-F52, L43-Y54, V80-V106, A133-A149 Rhodopsin-like GPCR superfamily PR00237: L36-K60, G31-F52, A27-V49, L91-F112, I35-I58, L85-W109, S123-A149 G-protein coupled receptors: DM00013 P51472 36-326: P2-C151 DM00013 S39028 36-326: P2-C151 DM00013 P32312 31-322: P2-C151 DM00013 P32310 35-326: P2-C151 Transmembrane domains: S33-I55, F187-V207	HMMER-PFAM MOTIFS ProfileScan BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DOMO
9	7477374CD1	332	S224 S228 S311 S6 T109 T90 Y99	N19 N4	7 transmembrane receptor (rhodopsin family): G49-Y131, V183-Y295 G-protein coupled receptor BL00237: W98-W137, F190-Y201, K235-D261, N287-G303 Rhodopsin-like GPCR superfamily PR00237: T67-F88, D112-G134, W182-F205, A240-I264, Y277-G303, I34-L58 G-protein coupled receptors: DM00013 S55549 13-327: M132-V310 DM00013 P32251 21-399: P26-A240 DM00013 I49480 45-449: Y27-A222 DM00013 P18825 45-452: Y27-I215	HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9					G-protein coupled receptor PD000009: K61-M132	BLAST- PRODOM
10	7479890CD1	948	S107 S144 S19 S488 S489 S666 S688 S843 S855 T309 Y456 Y504	N189	Signal peptide: M1-G24 Signal peptide: M1-A25 Transmembrane domains: A548-G571, V756-P783 Leucine rich repeats: R239-P262, L263-P286, K287-T309, S310-P333, R334-Q355, K356-S379, S380-H403, S404-M427, S144-P167, A168-T191, S192-H215, N216-G238 Leucine rich repeat N-terminal domain: A34-D65 Leucine zipper pattern: L57-L78 G-protein coupled receptor BL00237: R616-V655, P800-R816 Glycoprotein hormone receptor PR00373: F533-W550, F610-C623, C623-S637, W748-L766 G-protein coupled receptors: DM00013 P14763 407-693: P535-L819 DM00013 P35376 355-641: P535-D818 DM00013 P22888 352-638: P535-D818 DM00013 P35409 519-807: I545-L819 Orphan G-protein coupled receptor HG38: PD175529: H428-F568 PD169963: M259-P333 PD166277: A168-H215 G-PROTEIN COUPLED RECEPTORS: DM00013 P23270 18-311: L25-L306	BLAST- PRODOM SPScan HMMER HMMER HMMER-PFAM HMMER-PFAM HMMER-PFAM BLAST-DOMO
11	7482825CD1	315	S234 S54 S69 T10 T292 T6	N67 N8		BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11					RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L167-L246 G-protein coupled receptors proteins BL00237: T283-H299, A91-P130, L208-Y219, R236-A262 Olfactory receptor signature PR00245: M61-K82, F178-D192, F239-G254, F275-F286, T292-L306 Rhodopsin-like GPCR superfamily signature PR00237: L28-V52, M61-K82, F105-I127, T200-V223, K273-H299 G-protein coupled receptors signature g_protein_receptor.prf: F103-T149 Transmembrane domain: F27-V51, F102-D122, M143-A165, L201-T217 7 transmembrane receptor (rhodopsin family) 7tm_1: G43-Y291 G Protein Receptor: A111-I127	BLAST- PRODOM BLIMPS- BLOCKS BLIMPS- PRINTS BLIMPS- PRINTS PROFILESCAN HMMER HMMER-PFAM MOTIFS
12	7483087CD1	312	S152 S267 S268 S291 S67	N5 N93	G-PROTEIN COUPLED RECEPTORS: DM00013 P23265 17-306: E22-I305 DM00013 P23268 18-307: D20-I305 DM00013 P30955 18-305: D20-I305 DM00013 S29707 18-306: P21-L301 RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L166-I246 OLFACTORY RECEPTOR PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD149621: V247-R307	BLAST-DOMO BLAST- PRODOM BLAST- PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12					G-protein coupled receptors proteins BL00237: T282-L298, Q90-P129 Olfactory receptor signature PR00245: M59-K80, F177-D191, F238-G253, A274-L285, S291-I305 Transmembrane domain: L30-I46, Q100-M118, L143-M162, I197-F216 7 transmembrane receptor (rhodopsin family) 7tm_1: G41-Y290 G Protein Receptor: L110-I126 G-protein coupled receptors signature g_protein_receptor.prf: C102-V151 G-PROTEIN COUPLED RECEPTORS: DM00013 S297I10 15-301: F26-F300 DM00013 P23266 17-306: E20-L299 DM00013 S29708 18-306: E20-M296 DM00013 P23274 18-306: E20-M296 RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L164-I243 G-protein coupled receptors proteins BL00237: K88-P127, P277-K293 Rhodopsin-like GPCR superfamily signature PR00237: T24-N48, M57-K78, A102-I124, K138-L159, V197-L220, K267-K293 Olfactory receptor signature PR00245: M57-K78, F175-D189, V235-M250, T286-F300 Transmembrane domains: V28-I45, I204-N222	BLIMPS-BLOCKS BLIMPS-PRINTS HMMER HMMER-PFAM MOTIFS PROFILES SCAN BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS HMMER
13	7483134CD1	309	S227 S305 S65 T286 T52 T83	N6		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13					7 transmembrane receptor (rhodopsin family) 7tm_1: G39-Y285 G Protein Receptor: T108-I124 G-protein coupled receptors signature g_protein_receptor.prf: F100-A145 Signal cleavage: M1-G39 Transmembrane domains: L68-I86, M238-S256 7 transmembrane receptor (rhodopsin family): G81-Y292 G-protein coupled receptors signature: F142-G186 G-protein coupled receptor BL00237: K130-S169 (E-value<0.018) Olfactory receptor signature PR00245: M99-K120, F217-D231, F278-A293 Rhodopsin-like GPCR superfamily signature PR00237: F144-I166, M180-V201, V239-L262, A277-R301, F66-H90, M99-K120 RECEPTOR OLFATORY G PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: I206-L285 G-PROTEIN COUPLED RECEPTORS DM00013 P23266 17-306: L67-L307 DM00013 S51356 18-307: L67-S305 DM00013 S29707 18-306: L58-T308 DM00013 P23274 18-306: L58-T308 G-protein coupled receptors signature: S150-I166	HMMER_PFAM MOTIFS PROFILES SPSCAN HMMER HMMER_PFAM PROFILES BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS
14	7478550CD1	309	S107 T21 T26 T3 T48	N46 N61		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	7483142CD1	315	S191 S294 S70	N68 N8	Signal cleavage: M1-G54 Transmembrane domain: F31-L51 7 transmembrane receptor (rhodopsin family): W44-Y293 G-protein coupled receptor BL00237: K93-P132, W238-R264, I285-K301 Olfactory receptor signature PR00245: F180-D194, C241-G256, V277-L288, S294-L308, M62-N83 GPR orphan receptor signature PR00644: I52-Y63, S294-W305 G-protein coupled receptors signature: F105-V153 RECEPTOR OLFACTORY G PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY: PD000921: L169-M249 PD149621: V250-K311 G-PROTEIN COUPLED RECEPTORS: DM00013 P37067 17-306: T21-L304 DM00013 S29709 11-299: T21-L308 DM00013 P23266 17-306: P24-L308 DM00013 S51356 18-307: T21-L304 G-protein coupled receptors signature: S113-I129	SPSCAN HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS PROFILESCAN BLAST-PRODOR BLAST-DOMO MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	7483151CD1	307	S230 S261 S266 S291 T78	N107	Transmembrane domains: F28-A47, I196-I215 7 transmembrane receptor (rhodopsin family): G41-Y290 G-protein coupled receptor BL00237: H90-P129, I282-I298 G-protein coupled receptors signature: Y102-F150 Olfactory receptor signature PR00245: S291-M305, I59-K80, F177-N191, F238-G253, Y274-L285 Rhodopsin-like GPCR superfamily signature PR00237: I59-K80, F104-I126, I201-A224, K272-I298, P26-W50 RECEPTOR OLFACTORY G PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY: PD000921: L166-L246 PD149621: V248-M305 G-PROTEIN COUPLED RECEPTORS DM00013 P37067 17-306: L17-I304 DM00013 S51356 18-307: L17-R303 DM00013 S29709 11-299: T18-M305 DM00013 P23274 18-306: Q24-M305 G-protein coupled receptors signature: T110-I126	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
17	2536292CB1	2422	288-1079, 2271-2422, 105-150, 1835-1898	72051732V1	787	2422
				2536292F6 (BRAINOT18)	836	1680
				FL2536292_g6007890_000006_g499 5709	301	1175
				55062753H1	1	636
				55062751J1	480	2422
				55062763H1	648	2422
				72049620V1	1488	2422
				55062754J1	658	2422
				55094142J1	1	639
				FL7477708_g8217793_000019_g552 5078	338	1912
18	7477708CB1	1912	1798-1912, 1-1741, 1304-1481	1971428H1 (UCMCL5T01)	1028	1255
				GBI.g7547159_000017_000014.edi t	144	782
				GNN.g7547159_000014_004	517	1326
				GNN.g7106155_000002_002.edit	1	516
				71737648V1	1922	2535
20	644692CB1	3058	1-1633, 2354-3058, 2369-2504, 1-2201	55052474J1	370	1195
				71735994V1	1490	2079
				55067945J1	1	671
				71907834V1	2612	3058
				71736314V1	2293	2985
				6933242R8 (SINTTMR02)	910	1678
				g3738039	1088	1541
				FL3837054_g8575884_g1150834_1_2	849	1349
				GNN.g8575884_004	359	1420
				56005250J1	1649	1993
21	3837054CB1	1993	1-1063, 1540-1714, 1-87, 1543-1993, 447-829	3837054F6 (DENDTNT01)	1371	1935
				2814920H1 (OVARNOT10)	1	187

Table 4 (cont).

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
21				3406743H1 (PROSTUS08)	103	354
22	6157025CB1	1499	1-81, 248-1499, 850-920, 1265-1499	GNN.g5686520_000054_002.edit 6834181F6 (BRSTN0N02) FL6157025_g10567930_000001_g53 53887_1_1 7169316F8 (MCLRNOC01)	116 826 381	565 1499 1296
23	55012817CB1	2455	1397-1482, 318-704, 1893-2171, 1-1014, 1619-2182	4425914H1 (BRAPDIT01) 71691523V1 71691480V1 71691220V1 71690572V1 71691571V1 71691384V1 71688567V1	1 32 1746 1405 1 654 608 1351 2162	676 288 2401 2082 646 1377 1338 2058 2455
24	7475061CB1	2056	1-688, 1207-2056, 1736-1917, 1207-1680	72488727D1 72489896D1 72487236D1 72492196D1	841 1 1046 1200	1769 953 2056 2056
25	7477374CB1	999	1-138, 264-999	FL7477374_g9930948_000007_6739 496	1	999
26	7479890CB1	3429	1-2724, 2850-2900, 1-2300, 2367-2732	4021537F6 (BRAXNOT02) GBI.g9967464_000017_000007_000 001.edit 55017814H1 58013229J1 8042178J1 (OVARTUE01) GNN.g9967464_000015_002 71702678V1 6609077H2 (PLACFEC01) 7720776H1 (THYRDIE01) 7726057J1 (THYRDIE01)	1427 1 737 2532 2032 729 2160 366 880 1242	2110 588 1280 3429 2692 2847 2786 877 1291 1974

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
27	7482825CB1	948	1-87, 210-299, 923-948	GNN:g10045182_000002_006	1	948
38	7483087CB1	939	259-939, 1-86, 912-939	GNN:g6015288_000044_010	1	939
29	7483134CB1	930	1-327, 1-100, 889-930, 400-930	GNN:g7143464_000018_002	1	930
30	7478550CB1	1161	112-335, 116-358, 925-1161	7077972R8 (BRAUTDR04) 7077972F8 (BRAUTDR04)	1 416	926 1161
31	7483142CB1	948	1-26, 605-782, 913-948	FL7483142_g8086488_000017_g374 6443_1_1	1	948
32	7483151CB1	924	876-924	FL7483151_g9438337_000003_g151 4480_1_1-2	1	924

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
17	2536292CB1	BRAINOT18
19	7474823CB1	UCMCL5T01
20	644692CB1	SINVTMR02
21	3837054CB1	DENDTNT01
22	6157025CB1	MONOTXN05
23	55012817CB1	BRAPDIT01
26	7479890CB1	LJUNGNON07
30	7478550CB1	BRAUTDR04

Table 6

Library	Vector	Library Description
BRAINOT18	pINCY	Library was constructed using RNA isolated from left temporal lobe brain tissue removed from a 34-year-old Caucasian male during cerebral meninges lesion excision. Pathology for the associated tumor tissue indicated metastatic malignant melanoma. Neoplastic cells strongly expressed HMB-45. Patient history included malignant melanoma of skin of the trunk. Family history included liver cancer, acute myocardial infarction, atherosclerotic coronary artery disease, and cerebrovascular disease.
BRAPDIT01	pINCY	Library was constructed using RNA isolated from diseased pons tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Serology was negative. Patient history included Huntington's disease, emphysema, and tobacco abuse.
BRAUTDR04	PCDNA2.1	Library was constructed using RNA isolated from striatum, dorsal caudate nucleus, dorsal putamen, and ventral nucleus accumbens tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated no diagnostic abnormalities in the brain or intracranial vessels. There was mild meningeal fibrosis predominately over the convexities. Special stains showed no evidence of amyloid plaques or metastatic lesions. There were scattered axonal spheroids in the white matter of the cingulate cortex and thalamus. There were a few scattered neurofibrillary tangles in the entorhinal cortex and periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor, surrounded by foci of bile lakes beneath the hepatic surface scar. The liver had extensive surface scarring, congestion, cholestasis, hemorrhage, necrosis, and chronic inflammation. The patient presented with nausea, vomiting, dehydration, malnutrition, oliguria, and acute renal failure. Patient history included post-operative Budd-Chiari syndrome, biliary ascites, acute bilateral bronchopneumonia with microabscesses, hydrothorax, and bilateral leg pitting edema. Previous surgeries included cholecystectomy, liver resection, hysterectomy, bilateral salpingo-oophorectomy, and portocaval shunt. The patient was treated with a nasogastric feeding tube, biliary drainage stent, paracentesis, pleurodesis, and abdominal ultrasound. Patient medications included Ampicillin, niacin, furosemide, Aldactone, Benadryl, and morphine.

Table 6 (cont.)

Library	Vector	Library Description
DENDTNT01	pINCY	Library was constructed using RNA isolated from treated dendritic cells from peripheral blood.
LUNGNON07	pINCY	This normalized lung tissue library was constructed from 5.1 million independent clones from a lung tissue library. Starting RNA was made from RNA isolated from lung tissue. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
MONOTXN05	pINCY	This normalized treated monocyte cell tissue library was constructed from 1.03 million independent clones from a monocyte tissue library. Starting RNA was made from RNA isolated from treated monocytes from peripheral blood removed from a 42-year-old female. The cells were treated with interleukin-10 (IL-10) and lipopolysaccharide (LPS). The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
SINTTMR02	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 59-year-old male. Pathology for the matched tumor tissue indicated multiple (9) carcinoid tumors, grade 1, in the small bowel. The largest tumor was associated with a large mesenteric mass. Multiple convoluted segments of bowel were adhered to the tumor. A single (1 of 13) regional lymph node was positive for malignancy. The peritoneal biopsy indicated focal fat necrosis.
UCMCL5T01	PBLUESCRIPT	Library was constructed using RNA isolated from mononuclear cells obtained from the umbilical cord blood of 12 individuals. The cells were cultured for 12 days with IL-5 before RNA was obtained from the pooled lysates.

Table 7

Program	Description	Reference	Parameter Threshold
ABI/FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Atwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score \geq GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

1. An isolated polypeptide selected from the group consisting of:
 - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16,
 - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16,
 - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and
 - d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-16.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:17-32.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of producing a polypeptide of claim 1, the method comprising:
 - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
 - b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

5

12. An isolated polynucleotide selected from the group consisting of:

- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32,
- c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- e) an RNA equivalent of a)-d).

15

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
- b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

25

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

30

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- b) detecting the presence or absence of said amplified target polynucleotide or fragment

35

thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

5

18. A composition of claim 17, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.

19. A method for treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- 15
- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
 - b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

20

22. A method for treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- 25
- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
 - b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

30

25. A method for treating a disease or condition associated with overexpression of functional GCREC, comprising administering to a patient in need of such treatment a composition of claim 24.

35

26. A method of screening for a compound that specifically binds to the polypeptide of claim

1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- 5 b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- 10 a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- 15 c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- 25 b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

- 30 a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a
- 35 polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of GCREC in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')₂ fragment, or
- e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of GCREC in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of GCREC in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim

11, the method comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- 5 b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.

10 37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim

15 11, the method comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- 20 c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ
- 25 ID NO:1-16.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

30

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant

immunoglobulin library.

44. A method of detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16 in a sample, the method comprising:

- 5 a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16 in the sample.

10 45. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- 15 b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

20 47. A method of generating a transcript image of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- 25 c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is

completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A method of identifying a compound that modulates, mimics and/or blocks an olfactory and/or taste sensation, the method comprising:

- a) contacting the compound with an olfactory and/or taste receptor polypeptide selected from the group consisting of:
 - i) a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16,
 - ii) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, and
 - iii) an olfactory and/or taste receptor having an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16.
- b) identifying whether the compound specifically binds to and/or affects the activity of

said receptor polypeptide.

57. The method of claim 56, wherein said receptor polypeptide is expressed on the surface of a mammalian cell.

5

58. The method of claim 57, wherein said mammalian cell expresses a G-protein.

59. The method of claim 58, wherein said mammalian cell expresses a plurality of G-protein coupled receptors.

10

60. The method of claim 59, wherein said mammalian cell expresses another olfactory and/or taste receptor polypeptide.

15

61. The method of claim 56, wherein said receptor polypeptide is fused to another polypeptide.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

20

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

25

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

30

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.
72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.
- 5 73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
- 10 76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
- 15 78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.
79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.
80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.
- 20 81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.
82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.
- 25 83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.
84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.
85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.
- 30 86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.
87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

5 90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

10

93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

PI-0236 PCT

<110> INCYTE GENOMICS, INC.

BAUGHN, Mariah R.
GRAUL Richard C.
WALIA, Narinder K.
GANDHI, Ameena R.
HAFALIA, April J.A.
RAMKUMAR, Jayalaxmi
TRIBOULEY, Catherine M.
THORNTON, Michael
KALLICK, Deborah A.
YAO, Monique G.
ELLIOTT, Vicki S.
BURFORD, Neil
KHAN, Farrah A.
YUE, Henry
LU, Yan
ARVIZU, Chandra
ROOPA, Reddy
NGUYEN, Danniel B.
LEE, Ernestine A.
LU, Dyung Aina M.
ISON, Craig H.
WALSH, Roderick T.
POLICKY, Jennifer L.

<120> G-PROTEIN COUPLED RECEPTORS

<130> PI-0236 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/236,546; 60/242,223; 60/245,900; 60/247,587; 60/249,343; 60/240,589;
60/242,322; 60/245,855

<151> 2000-09-29; 2000-10-20; 2000-11-03; 2000-11-09; 2000-11-15; 2000-10-13;
2000-10-20; 2000-11-03

<160> 32

<170> PERL Program

<210> 1

<211> 217

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2536292CD1

<400> 1

Met Lys Ser Phe Leu Pro Gly Thr Cys Ile Leu Leu Cys Ser Ala

1

5

10

15

PI-0236 PCT

```

Phe Asn Leu Met Phe Phe Ser Leu Phe Arg Leu Lys Tyr Asn Ile
      20      25      30
Cys Ile Ile Leu Arg Ala Cys Asn Thr Met Leu Ser Ser Asn Thr
      35      40      45
Ile Met Glu Ile Phe Phe Leu Ser His Ile Asp Ile Gly Ile Trp
      50      55      60
Arg Asn Leu Leu Leu Leu Leu Met Pro Ile Tyr Thr Phe Leu Ile
      65      70      75
Cys Pro Gln Gln Lys Lys Pro Met Gly Leu Leu Phe Leu His Leu
      80      85      90
Ser Val Ala Asn Thr Met Thr Leu Leu Arg Lys Val Ile Pro Leu
      95     100     105
Ala Val Lys Ser Phe Asn Thr Lys Asn Leu Leu Asn Tyr Thr Gly
     110     115     120
Cys Arg Glu Phe Glu Phe Leu Tyr Arg Val Ser Trp Gly Leu Pro
     125     130     135
Leu Cys Thr Thr Tyr Leu Leu Ser Met Val Gln Ala Leu Arg Gly
     140     145     150
Ser Pro Ser Lys Ser Arg Val Ile Asn Ser Leu Ile Tyr Ile Lys
     155     160     165
Leu Val Pro Phe Val Asp Thr Ile Lys Tyr Gly Ser Val Thr Lys
     170     175     180
Asn Leu Ser Ile Lys Met Cys Leu Ala Thr Pro His Met Gly Asn
     185     190     195
Thr Ile Ala Val Ser His Thr Ser Val Ile Thr Phe Gln Asp Leu
     200     205     210
Ile Phe Leu Val Leu Met Ser
      215

```

<210> 2

<211> 578

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7477708CD1

<400> 2

```

Met Gly Phe Ser Cys Arg Gln Lys Thr Trp His Lys Ile Thr Asp
  1      5      10      15
Thr Cys Arg Thr Leu Asn Ala Leu Asn Ile Phe Glu Glu Asp Ser
     20      25      30
Arg Leu Val Gln Pro Phe Glu Asp Asn Ile Lys Ile Ser Val Tyr
     35      40      45
Thr Gly Lys Ser Glu Thr Ile Thr Asp Met Leu Leu Gln Lys Cys
     50      55      60
Pro Thr Asp Leu Ser Cys Val Ile Arg Asn Ile Gln Gln Ser Pro
     65      70      75
Trp Ile Pro Gly Asn Ile Ala Val Ile Val Gln Leu Leu His Asn
     80      85      90
Ile Ser Thr Ala Ile Trp Thr Gly Val Asp Glu Ala Lys Met Gln

```

PI-0236 PCT

	95	100	105
Ser Tyr Ser Thr	Ile Ala Asn His Ile	Leu Asn Ser Lys Ser	Ile
	110	115	120
Ser Asn Trp Thr	Phe Ile Pro Asp Arg	Asn Ser Ser Tyr Ile	Leu
	125	130	135
Leu His Ser Val	Asn Ser Phe Ala Arg	Arg Leu Phe Ile Asp	Asn
	140	145	150
Ile Pro Val Asp	Ile Ser Asp Val Phe	Ile His Thr Met Gly	Thr
	155	160	165
Thr Ile Ser Gly	Asp Asn Ile Gly Lys	Asn Phe Thr Phe Ser	Met
	170	175	180
Arg Ile Asn Asp	Thr Ser Asn Glu Val	Thr Gly Arg Val Leu	Ile
	185	190	195
Ser Arg Asp Glu	Leu Arg Lys Val Pro	Ser Pro Ser Gln Val	Ile
	200	205	210
Ser Ile Ala Phe	Pro Thr Ile Gly Ala	Ile Leu Glu Ala Ser	Leu
	215	220	225
Leu Glu Asn Val	Thr Val Asn Gly Leu	Val Leu Ser Ala Ile	Leu
	230	235	240
Pro Lys Glu Leu	Lys Arg Ile Ser Leu	Ile Phe Glu Lys Ile	Ser
	245	250	255
Lys Ser Glu Glu	Arg Arg Thr Gln Cys	Val Gly Trp His Ser	Val
	260	265	270
Glu Asn Arg Trp	Asp Gln Gln Ala Cys	Lys Met Ile Gln Glu	Asn
	275	280	285
Ser Gln Gln Ala	Val Cys Lys Cys Arg	Pro Ser Lys Leu Phe	Thr
	290	295	300
Ser Phe Ser Ile	Leu Met Ser Pro His	Ile Leu Glu Ser Leu	Ile
	305	310	315
Leu Thr Tyr Ile	Thr Tyr Val Gly Leu	Gly Ile Ser Ile Cys	Ser
	320	325	330
Leu Ile Leu Cys	Leu Ser Ile Glu Val	Leu Val Trp Ser Gln	Val
	335	340	345
Thr Lys Thr Glu	Ile Thr Tyr Leu Arg	His Val Cys Ile Val	Asn
	350	355	360
Ile Ala Ala Thr	Leu Leu Met Ala Asp	Val Trp Phe Ile Val	Ala
	365	370	375
Ser Phe Leu Ser	Gly Pro Ile Thr His	His Lys Gly Cys Val	Ala
	380	385	390
Ala Thr Phe Phe	Val His Phe Phe Tyr	Leu Ser Val Phe Phe	Trp
	395	400	405
Met Leu Ala Lys	Ala Leu Leu Ile Leu	Tyr Gly Ile Met Ile	Val
	410	415	420
Phe His Thr Leu	Pro Lys Ser Val Leu	Val Ala Ser Leu Phe	Ser
	425	430	435
Val Gly Tyr Gly	Cys Pro Leu Ala Ile	Ala Ala Ile Thr Val	Ala
	440	445	450
Ala Thr Glu Pro	Gly Lys Gly Tyr Leu	Arg Pro Glu Ile Cys	Trp
	455	460	465
Leu Asn Trp Asp	Met Thr Lys Ala Leu	Leu Ala Phe Val Ile	Pro
	470	475	480
Ala Leu Ala Ile	Val Val Val Asn Leu	Ile Thr Val Thr Leu	Val

PI-0236 PCT

	485	490	495
Ile Val Lys Thr	Gln Arg Ala Ala Ile	Gly Asn Ser Met Phe	Gln
	500	505	510
Glu Val Arg Ala	Ile Val Arg Ile Ser	Lys Asn Ile Ala Ile	Leu
	515	520	525
Thr Pro Leu Leu	Gly Leu Thr Trp Gly	Phe Gly Val Ala Thr	Val
	530	535	540
Ile Asp Asp Arg	Ser Leu Ala Phe His	Ile Ile Phe Ser Leu	Leu
	545	550	555
Asn Ala Phe Gln	Val Ser Pro Asp Ala	Ser Asp Gln Val Gln	Ser
	560	565	570
Glu Arg Ile His	Glu Asp Val Leu		
	575		

<210> 3

<211> 441

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7474823CD1

<400> 3

Met Val Leu Gly Lys Asn Val Ser Met Ser Gly Pro Arg Pro Ala		
1	5	10
Ser Trp Gln Ser His Pro Gln Gly Leu Glu Leu Val Phe Gly Lys		
	20	25
Trp Pro Cys Arg Cys Ser Tyr Ser Ala Val Leu Val Ile Ser Ser		
	35	40
Ile Ser Ser Ser Gly Ala Gly Asp Ile Pro Asp Gln Asp Ser Gly		
	50	55
Gln Tyr Trp Phe Leu Met Arg Ala Val Phe Leu Ala Cys Arg Arg		
	65	70
Leu Pro Ser Thr Cys Val Leu Lys Arg Pro Phe Ser Glu Cys Ala		
	80	85
Gln Arg Glu Arg Thr Asn Leu Val Leu Met Lys Lys Trp Glu Phe		
	95	100
Leu Glu Val Pro Asp Thr Phe Glu Val Thr Gln Gln Ser Val Ile		
	110	115
Ser Ile Pro Leu Tyr Ile Pro His Thr Leu Phe Glu Trp Asp Phe		
	125	130
Gly Lys Glu Ile Cys Val Phe Trp Leu Thr Thr Asp Tyr Leu Leu		
	140	145
Cys Thr Ala Ser Val Tyr Asn Ile Val Leu Ile Ser Tyr Asp Arg		
	155	160
Tyr Leu Ser Val Ser Asn Ala Val Ser Tyr Arg Thr Gln His Thr		
	170	175
Gly Val Leu Lys Ile Val Thr Leu Met Val Ala Val Trp Val Leu		
	185	190
Ala Phe Leu Val Asn Gly Pro Met Ile Leu Val Ser Glu Ser Trp		
	200	205
		210

PI-0236 PCT

Lys Asp Glu Gly Ser Glu Cys Glu Pro Gly Phe Phe Ser Glu Trp
 215 220 225
 Tyr Ile Leu Ala Ile Thr Ser Phe Leu Glu Phe Val Ile Pro Val
 230 235 240
 Ile Leu Val Ala Tyr Phe Asn Met Asn Ile Tyr Trp Ser Leu Trp
 245 250 255
 Lys Arg Asp His Leu Arg Leu Gly His Pro Lys Gly Trp Gly Gln
 260 265 270
 Leu Val Leu Arg Leu Pro His Gly Val Glu Gly Gln Pro Trp Arg
 275 280 285
 Leu Gln Leu Val Pro Arg Met Gly Tyr Ile Glu Val Gly Gly Leu
 290 295 300
 Leu Cys Thr Ala Ala Gly Glu Met Ser Thr His Ala Arg Ser Ala
 305 310 315
 Lys Leu Leu Ser Thr Gly Ser Glu Asn Asp Thr Leu Pro Val Pro
 320 325 330
 Ser Leu Ala Ser Arg Ser Leu Cys Pro Ser Val Leu Ser Leu Gly
 335 340 345
 Ser Phe Pro Ser Cys Gln Ser Cys Leu Ser Asp Gln Met Ser Gln
 350 355 360
 Cys Asp Thr Glu Pro Glu Arg Lys Ser Phe Leu Ser Met Met Gln
 365 370 375
 Gly Thr Gln His Phe Asp Asn Pro Asp Gly Met Trp Ser Ser His
 380 385 390
 Gly Arg Asn Val Ser Ser Gly Gly Leu His Asn His Cys Ile Leu
 395 400 405
 Gln Met Gly Thr Gly Ser Ala Gly Ala Ser His Pro Glu Gly Pro
 410 415 420
 Arg Gly Gly Gln Gly Gln Val Thr Thr Arg Ala Thr Thr Gln Lys
 425 430 435
 Arg Val Ala Ala Ser Gly
 440

<210> 4

<211> 797

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 644692CD1

<400> 4

Met Ala Ser Cys Arg Ala Trp Asn Leu Arg Val Leu Val Ala Val
 1 5 10 15
 Val Cys Gly Leu Leu Thr Gly Ile Ile Leu Gly Leu Gly Ile Trp
 20 25 30
 Arg Ile Val Ile Arg Ile Gln Arg Gly Lys Ser Thr Ser Ser Ser
 35 40 45
 Ser Thr Pro Thr Glu Phe Cys Arg Asn Gly Gly Thr Trp Glu Asn
 50 55 60
 Gly Arg Cys Ile Cys Thr Glu Glu Trp Lys Gly Leu Arg Cys Thr

PI-0236 PCT

				65					70					75
Ile	Ala	Asn	Phe	Cys	Glu	Asn	Ser	Thr	Tyr	Met	Gly	Phe	Thr	Phe
				80					85					90
Ala	Arg	Ile	Pro	Val	Gly	Arg	Tyr	Gly	Pro	Ser	Leu	Gln	Thr	Cys
				95					100					105
Gly	Lys	Asp	Thr	Pro	Asn	Ala	Gly	Asn	Pro	Met	Ala	Val	Arg	Leu
				110					115					120
Cys	Ser	Leu	Ser	Leu	Tyr	Gly	Glu	Ile	Glu	Leu	Gln	Lys	Val	Thr
				125					130					135
Ile	Gly	Asn	Cys	Asn	Glu	Asn	Leu	Glu	Thr	Leu	Glu	Lys	Gln	Val
				140					145					150
Lys	Asp	Val	Thr	Ala	Pro	Leu	Asn	Asn	Ile	Ser	Ser	Glu	Val	Gln
				155					160					165
Ile	Leu	Thr	Ser	Asp	Ala	Asn	Lys	Leu	Thr	Ala	Glu	Asn	Ile	Thr
				170					175					180
Ser	Ala	Thr	Arg	Val	Val	Gly	Gln	Ile	Phe	Asn	Thr	Ser	Arg	Asn
				185					190					195
Ala	Ser	Pro	Glu	Ala	Lys	Lys	Val	Ala	Ile	Val	Thr	Val	Ser	Gln
				200					205					210
Leu	Leu	Asp	Ala	Ser	Glu	Asp	Ala	Phe	Gln	Arg	Val	Ala	Ala	Thr
				215					220					225
Ala	Asn	Asp	Asp	Ala	Leu	Thr	Thr	Leu	Ile	Glu	Gln	Met	Glu	Thr
				230					235					240
Tyr	Ser	Leu	Ser	Leu	Gly	Asn	Gln	Ser	Val	Val	Glu	Pro	Asn	Ile
				245					250					255
Ala	Ile	Gln	Ser	Ala	Asn	Phe	Ser	Ser	Glu	Asn	Ala	Val	Gly	Pro
				260					265					270
Ser	Asn	Val	Arg	Phe	Ser	Val	Gln	Lys	Gly	Ala	Ser	Ser	Ser	Leu
				275					280					285
Val	Ser	Ser	Ser	Thr	Phe	Ile	His	Thr	Asn	Val	Asp	Gly	Leu	Asn
				290					295					300
Pro	Asp	Ala	Gln	Thr	Glu	Leu	Gln	Val	Leu	Leu	Asn	Met	Thr	Lys
				305					310					315
Asn	Tyr	Thr	Lys	Thr	Cys	Gly	Phe	Val	Val	Tyr	Gln	Asn	Asp	Lys
				320					325					330
Leu	Phe	Gln	Ser	Lys	Thr	Phe	Thr	Ala	Lys	Ser	Asp	Phe	Ser	Gln
				335					340					345
Lys	Ile	Ile	Ser	Ser	Lys	Thr	Asp	Glu	Asn	Glu	Gln	Asp	Gln	Ser
				350					355					360
Ala	Ser	Val	Asp	Met	Val	Phe	Ser	Pro	Lys	Tyr	Asn	Gln	Lys	Glu
				365					370					375
Phe	Gln	Leu	Tyr	Ser	Tyr	Ala	Cys	Val	Tyr	Trp	Asn	Leu	Ser	Ala
				380					385					390
Lys	Asp	Trp	Asp	Thr	Tyr	Gly	Cys	Gln	Lys	Asp	Lys	Gly	Thr	Asp
				395					400					405
Gly	Phe	Leu	Arg	Cys	Arg	Cys	Asn	His	Thr	Thr	Asn	Phe	Ala	Val
				410					415					420
Leu	Met	Thr	Phe	Lys	Lys	Asp	Tyr	Gln	Tyr	Pro	Lys	Ser	Leu	Asp
				425					430					435
Ile	Leu	Ser	Asn	Val	Gly	Cys	Ala	Leu	Ser	Val	Thr	Gly	Leu	Ala
				440					445					450
Leu	Thr	Val	Ile	Phe	Gln	Ile	Val	Thr	Arg	Lys	Val	Arg	Lys	Thr

PI-0236 PCT

	455	460	465
Ser Val Thr Trp Val Leu Val Asn Leu Cys Ile Ser Met Leu Ile			
	470	475	480
Phe Asn Leu Leu Phe Val Phe Gly Ile Glu Asn Ser Asn Lys Asn			
	485	490	495
Leu Gln Thr Ser Asp Gly Asp Ile Asn Asn Ile Asp Phe Asp Asn			
	500	505	510
Asn Asp Ile Pro Arg Thr Asp Thr Ile Asn Ile Pro Asn Pro Met			
	515	520	525
Cys Thr Ala Ile Ala Ala Leu Leu His Tyr Phe Leu Leu Val Thr			
	530	535	540
Phe Thr Trp Asn Ala Leu Ser Ala Ala Gln Leu Tyr Tyr Leu Leu			
	545	550	555
Ile Arg Thr Met Lys Pro Leu Pro Arg His Phe Ile Leu Phe Ile			
	560	565	570
Ser Leu Ile Gly Trp Gly Val Pro Ala Ile Val Val Ala Ile Thr			
	575	580	585
Val Gly Val Ile Tyr Ser Gln Asn Gly Asn Asn Pro Gln Trp Glu			
	590	595	600
Leu Asp Tyr Arg Gln Glu Lys Ile Cys Trp Leu Ala Ile Pro Glu			
	605	610	615
Pro Asn Gly Val Ile Lys Ser Pro Leu Leu Trp Ser Phe Ile Val			
	620	625	630
Pro Val Thr Ile Ile Leu Ile Ser Asn Val Val Met Phe Ile Thr			
	635	640	645
Ile Ser Ile Lys Val Leu Trp Lys Asn Asn Gln Asn Leu Thr Ser			
	650	655	660
Thr Lys Lys Val Ser Ser Met Lys Lys Ile Val Ser Thr Leu Ser			
	665	670	675
Val Ala Val Val Phe Gly Ile Thr Trp Ile Leu Ala Tyr Leu Met			
	680	685	690
Leu Val Asn Asp Asp Ser Ile Arg Ile Val Phe Ser Tyr Ile Phe			
	695	700	705
Cys Leu Phe Asn Thr Thr Gln Gly Leu Gln Ile Phe Ile Leu Tyr			
	710	715	720
Thr Val Arg Thr Lys Val Phe Gln Ser Glu Ala Ser Lys Val Leu			
	725	730	735
Met Leu Leu Ser Ser Ile Gly Arg Arg Lys Ser Leu Pro Ser Val			
	740	745	750
Thr Arg Pro Arg Leu Arg Val Lys Met Tyr Asn Phe Leu Arg Ser			
	755	760	765
Leu Pro Thr Leu His Glu Arg Phe Arg Leu Leu Glu Thr Ser Pro			
	770	775	780
Ser Thr Glu Glu Ile Thr Leu Ser Glu Ser Asp Asn Ala Lys Glu			
	785	790	795

Ser Ile

<210> 5

<211> 434

<212> PRT

<213> Homo sapiens

PI-0236 PCT

<220>

<221> misc_feature

<223> Incyte ID No: 3837054CD1

<400> 5

Met	Glu	Asp	Leu	Phe	Ser	Pro	Ser	Ile	Leu	Pro	Pro	Ala	Pro	Asn
1				5					10					15
Ile	Ser	Val	Pro	Ile	Leu	Leu	Gly	Trp	Gly	Leu	Asn	Leu	Thr	Leu
				20					25					30
Gly	Gln	Gly	Ala	Pro	Ala	Ser	Gly	Pro	Pro	Ser	Arg	Arg	Val	Arg
				35					40					45
Leu	Val	Phe	Leu	Gly	Val	Ile	Leu	Val	Val	Ala	Val	Ala	Gly	Asn
				50					55					60
Thr	Thr	Val	Leu	Cys	Arg	Leu	Cys	Gly	Gly	Gly	Gly	Pro	Trp	Ala
				65					70					75
Gly	Pro	Lys	Arg	Arg	Lys	Met	Asp	Phe	Leu	Leu	Val	Gln	Leu	Ala
				80					85					90
Leu	Ala	Asp	Leu	Tyr	Ala	Cys	Gly	Gly	Thr	Ala	Leu	Ser	Gln	Leu
				95					100					105
Ala	Trp	Glu	Leu	Leu	Gly	Glu	Pro	Arg	Ala	Ala	Thr	Gly	Asp	Leu
				110					115					120
Ala	Cys	Arg	Phe	Leu	Gln	Leu	Leu	Gln	Ala	Ser	Gly	Arg	Gly	Ala
				125					130					135
Ser	Ala	His	Leu	Val	Val	Leu	Ile	Ala	Leu	Glu	Arg	Arg	Arg	Ala
				140					145					150
Val	Arg	Leu	Pro	His	Gly	Arg	Pro	Leu	Pro	Ala	Arg	Ala	Leu	Ala
				155					160					165
Ala	Leu	Gly	Trp	Leu	Leu	Ala	Leu	Leu	Leu	Ala	Leu	Pro	Pro	Ala
				170					175					180
Phe	Val	Val	Arg	Gly	Asp	Ser	Pro	Ser	Pro	Leu	Pro	Pro	Pro	Pro
				185					190					195
Pro	Pro	Thr	Ser	Leu	Gln	Pro	Gly	Ala	Pro	Pro	Ala	Ala	Arg	Ala
				200					205					210
Trp	Pro	Gly	Glu	Arg	Arg	Cys	His	Gly	Ile	Phe	Ala	Pro	Leu	Pro
				215					220					225
Arg	Trp	His	Leu	Gln	Val	Tyr	Ala	Phe	Tyr	Glu	Ala	Val	Ala	Gly
				230					235					240
Phe	Val	Ala	Pro	Val	Thr	Val	Leu	Gly	Val	Ala	Cys	Gly	His	Leu
				245					250					255
Leu	Ser	Val	Trp	Trp	Arg	His	Arg	Pro	Gln	Ala	Pro	Ala	Ala	Ala
				260					265					270
Ala	Pro	Trp	Ser	Ala	Ser	Pro	Gly	Arg	Ala	Pro	Ala	Pro	Ser	Ala
				275					280					285
Leu	Pro	Arg	Ala	Lys	Val	Gln	Ser	Leu	Lys	Met	Ser	Leu	Leu	Leu
				290					295					300
Ala	Leu	Leu	Phe	Val	Gly	Cys	Glu	Leu	Pro	Tyr	Phe	Ala	Ala	Arg
				305					310					315
Leu	Ala	Ala	Ala	Trp	Ser	Ser	Gly	Pro	Ala	Gly	Asp	Trp	Glu	Gly
				320					325					330
Glu	Gly	Leu	Ser	Ala	Ala	Leu	Arg	Val	Val	Ala	Met	Ala	Asn	Ser
				335					340					345

PI-0236 PCT

Ala Leu Asn Pro Phe Val Tyr Leu Phe Phe Gln Ala Gly Asp Cys
 350 355 360
 Arg Leu Arg Arg Gln Leu Arg Lys Arg Leu Gly Ser Leu Cys Cys
 365 370 375
 Ala Pro Gln Gly Gly Ala Glu Asp Glu Glu Gly Pro Arg Gly His
 380 385 390
 Gln Ala Leu Tyr Arg Gln Arg Trp Pro His Pro His Tyr His His
 395 400 405
 Ala Arg Arg Glu Pro Leu Asp Glu Gly Gly Leu Arg Pro Pro Pro
 410 415 420
 Pro Arg Pro Arg Pro Leu Pro Cys Ser Cys Glu Ser Ala Phe
 425 430

<210> 6

<211> 339

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6157025CD1

<400> 6

Met Pro Gly His Asn Thr Ser Arg Asn Ser Ser Cys Asp Pro Ile
 1 5 10 15
 Val Thr Pro His Leu Ile Ser Leu Tyr Phe Ile Val Leu Ile Gly
 20 25 30
 Gly Leu Val Gly Val Ile Ser Ile Leu Phe Leu Leu Val Lys Met
 35 40 45
 Asn Thr Arg Ser Val Thr Thr Met Ala Val Ile Asn Leu Val Val
 50 55 60
 Val His Ser Val Phe Leu Leu Thr Val Pro Phe Arg Leu Thr Tyr
 65 70 75
 Leu Ile Lys Lys Thr Trp Met Phe Gly Leu Pro Phe Cys Lys Phe
 80 85 90
 Val Ser Ala Met Leu His Ile His Met Tyr Leu Thr Phe Leu Phe
 95 100 105
 Tyr Val Val Ile Leu Val Thr Arg Tyr Leu Ile Phe Phe Lys Cys
 110 115 120
 Lys Asp Lys Val Glu Phe Tyr Arg Lys Leu His Ala Val Ala Ala
 125 130 135
 Ser Ala Gly Met Trp Thr Leu Val Ile Val Ile Val Val Pro Leu
 140 145 150
 Val Val Ser Arg Tyr Gly Ile His Glu Glu Tyr Asn Glu Glu His
 155 160 165
 Cys Phe Lys Phe His Lys Glu Leu Ala Tyr Thr Tyr Val Lys Ile
 170 175 180
 Ile Asn Tyr Met Ile Val Ile Phe Val Ile Ala Val Ala Val Ile
 185 190 195
 Leu Leu Val Phe Gln Val Phe Ile Ile Met Leu Met Val Gln Lys
 200 205 210
 Leu Arg His Ser Leu Leu Ser His Gln Glu Phe Trp Ala Gln Leu

PI-0236 PCT

	215		220		225
Lys Asn Leu Phe Phe Ile Gly Val Ile Leu Val Cys Phe Leu Pro					
	230		235		240
Tyr Gln Phe Phe Arg Ile Tyr Tyr Leu Asn Val Val Thr His Ser					
	245		250		255
Asn Ala Cys Asn Ser Lys Val Ala Phe Tyr Asn Glu Ile Phe Leu					
	260		265		270
Ser Val Thr Ala Ile Ser Cys Tyr Asp Leu Leu Leu Phe Val Phe					
	275		280		285
Gly Gly Ser His Trp Phe Lys Gln Lys Ile Met Ala Tyr Gly Ile					
	290		295		300
Val Phe Val Pro Leu Ala Thr Asn Tyr Ser Ile His Ile Cys Phe					
	305		310		315
Leu Tyr Ile Gly Asn Lys Asn Gly Tyr Arg Gly Gly Lys Asn Gly					
	320		325		330
Ile Ser Leu Leu Asp Gln Ser Met Pro					
	335				

<210> 7

<211> 549

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 55012817CD1

<400> 7

Met Ala Thr Pro Arg Gly Leu Gly Ala Leu Leu Leu Leu Leu Leu					
1	5		10		15
Leu Pro Thr Ser Gly Gln Glu Lys Pro Thr Glu Gly Pro Arg Asn					
	20		25		30
Thr Cys Leu Gly Ser Asn Asn Met Tyr Asp Ile Phe Asn Leu Asn					
	35		40		45
Asp Lys Ala Leu Cys Phe Thr Lys Cys Arg Gln Ser Gly Ser Asp					
	50		55		60
Ser Cys Asn Val Glu Asn Leu Gln Arg Tyr Trp Leu Asn Tyr Glu					
	65		70		75
Ala His Leu Met Lys Glu Gly Leu Thr Gln Lys Val Asn Thr Pro					
	80		85		90
Phe Leu Lys Ala Leu Val Gln Asn Leu Ser Thr Asn Thr Ala Glu					
	95		100		105
Asp Phe Tyr Phe Ser Leu Glu Pro Ser Gln Val Pro Arg Gln Val					
	110		115		120
Met Lys Asp Glu Asp Lys Pro Pro Asp Arg Val Arg Leu Pro Lys					
	125		130		135
Ser Leu Phe Arg Ser Leu Pro Gly Asn Arg Ser Val Val Arg Leu					
	140		145		150
Ala Val Thr Ile Leu Asp Ile Gly Pro Gly Thr Leu Phe Lys Gly					
	155		160		165
Pro Arg Leu Gly Leu Gly Asp Gly Ser Gly Val Leu Asn Asn Arg					
	170		175		180

PI-0236 PCT

Leu Val Gly Leu Ser Val Gly Gln Met His Val Thr Lys Leu Ala	185	190	195
Glu Pro Leu Glu Ile Val Phe Ser His Gln Arg Pro Pro Pro Asn	200	205	210
Met Thr Leu Thr Cys Val Phe Trp Asp Val Thr Lys Gly Thr Thr	215	220	225
Gly Asp Trp Ser Ser Glu Gly Cys Ser Thr Glu Val Arg Pro Glu	230	235	240
Gly Thr Val Cys Cys Cys Asp His Leu Thr Phe Phe Ala Leu Leu	245	250	255
Leu Arg Pro Thr Leu Asp Gln Ser Thr Val His Ile Leu Thr Arg	260	265	270
Ile Ser Gln Ala Gly Cys Gly Val Ser Met Ile Phe Leu Ala Phe	275	280	285
Thr Ile Ile Leu Tyr Ala Phe Leu Arg Leu Ser Arg Glu Arg Phe	290	295	300
Lys Ser Glu Asp Ala Pro Lys Ile His Val Ala Leu Gly Gly Ser	305	310	315
Leu Phe Leu Leu Asn Leu Ala Phe Leu Val Asn Val Gly Ser Gly	320	325	330
Ser Lys Gly Ser Asp Ala Ala Cys Trp Ala Arg Gly Ala Val Phe	335	340	345
His Tyr Phe Leu Leu Cys Ala Phe Thr Trp Met Gly Leu Glu Ala	350	355	360
Phe His Leu Tyr Leu Leu Ala Val Arg Val Phe Asn Thr Tyr Phe	365	370	375
Gly His Tyr Phe Leu Lys Leu Ser Leu Val Gly Trp Gly Leu Pro	380	385	390
Ala Leu Met Val Ile Gly Thr Gly Ser Ala Asn Ser Tyr Gly Leu	395	400	405
Tyr Thr Ile Arg Asp Arg Glu Asn Arg Thr Ser Leu Glu Leu Cys	410	415	420
Trp Phe Arg Glu Gly Thr Thr Met Tyr Ala Leu Tyr Ile Thr Val	425	430	435
His Gly Tyr Phe Leu Ile Thr Phe Leu Phe Gly Met Val Val Leu	440	445	450
Ala Leu Val Val Trp Lys Ile Phe Thr Leu Ser Arg Ala Thr Ala	455	460	465
Val Lys Glu Arg Gly Lys Asn Arg Lys Lys Val Leu Thr Leu Leu	470	475	480
Gly Leu Ser Ser Leu Val Gly Val Thr Trp Gly Leu Ala Ile Phe	485	490	495
Thr Pro Leu Gly Leu Ser Thr Val Tyr Ile Phe Ala Leu Phe Asn	500	505	510
Ser Leu Gln Gly Val Phe Ile Cys Cys Trp Phe Thr Ile Leu Tyr	515	520	525
Leu Pro Ser Gln Ser Thr Thr Val Ser Ser Ser Thr Ala Arg Leu	530	535	540
Asp Gln Ala His Ser Ala Ser Gln Glu	545		

<210> 8

PI-0236 PCT

<211> 188

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7475061CD1

<400> 8

```

Met Pro Leu Val Gly Leu Gly Asp Tyr Val Pro Glu Pro Phe Gly
  1              5              10              15
Thr Ser Cys Thr Leu Asp Trp Trp Leu Ala Gln Ala Ser Val Gly
              20              25              30
Gly Gln Val Phe Ile Leu Asn Ile Leu Phe Phe Cys Leu Leu Leu
              35              40              45
Pro Thr Ala Val Ile Val Phe Ser Tyr Val Lys Ile Ile Ala Lys
              50              55              60
Val Lys Ser Ser Ser Lys Glu Val Ala His Phe Asp Ser Arg Ile
              65              70              75
His Ser Ser His Val Leu Glu Met Lys Leu Thr Lys Val Ala Met
              80              85              90
Leu Ile Cys Ala Gly Phe Leu Ile Ala Trp Ile Pro Tyr Ala Val
              95              100             105
Val Ser Val Trp Ser Ala Phe Gly Arg Pro Asp Ser Ile Pro Ile
              110             115             120
Gln Leu Ser Val Val Pro Thr Leu Leu Ala Lys Ser Ala Ala Met
              125             130             135
Tyr Asn Pro Ile Ile Tyr Gln Val Ile Asp Tyr Lys Phe Ala Cys
              140             145             150
Cys Gln Thr Gly Gly Leu Lys Ala Thr Lys Lys Lys Ser Leu Glu
              155             160             165
Gly Phe Arg Leu His Thr Val Thr Thr Val Arg Lys Ser Ser Ala
              170             175             180
Val Leu Glu Ile His Glu Glu Val
              185

```

<210> 9

<211> 332

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7477374CD1

<400> 9

```

Met Val Asn Asn Phe Ser Gln Ala Glu Ala Val Glu Leu Cys Tyr
  1              5              10              15
Lys Asn Val Asn Glu Ser Cys Ile Lys Thr Pro Tyr Ser Pro Gly
              20              25              30
Pro Arg Ser Ile Leu Tyr Ala Val Leu Gly Phe Gly Ala Val Leu
              35              40              45

```

PI-0236 PCT

Ala Ala Phe Gly Asn Leu Leu Val Met Ile Ala Ile Leu His Phe
 50 55 60
 Lys Gln Leu His Thr Pro Thr Asn Phe Leu Ile Ala Ser Leu Ala
 65 70 75
 Cys Ala Asp Phe Leu Val Gly Val Thr Val Met Pro Phe Ser Thr
 80 85 90
 Val Arg Ser Val Glu Ser Cys Trp Tyr Phe Gly Asp Ser Tyr Cys
 95 100 105
 Lys Phe His Thr Cys Phe Asp Thr Ser Phe Cys Phe Ala Ser Leu
 110 115 120
 Phe His Leu Cys Cys Ile Ser Val Asp Arg Tyr Met Leu Gly Tyr
 125 130 135
 Ala Trp Phe Phe Pro Gly Phe Phe Ser Val Thr Tyr Ser Phe Ser
 140 145 150
 Ile Phe Asn Thr Gly Ala Asn Glu Glu Gly Ile Glu Glu Leu Val
 155 160 165
 Val Ala Leu Thr Cys Val Gly Gly Cys Gln Ala Pro Leu Asn Gln
 170 175 180
 Asn Trp Val Leu Leu Cys Phe Leu Leu Phe Phe Ile Pro Asn Val
 185 190 195
 Ala Met Val Phe Ile Tyr Ser Lys Ile Phe Leu Val Ala Lys His
 200 205 210
 Gln Ala Arg Lys Ile Glu Ser Thr Ala Ser Gln Ala Gln Ser Ser
 215 220 225
 Ser Glu Ser Tyr Lys Glu Arg Val Ala Lys Arg Glu Arg Lys Ala
 230 235 240
 Ala Lys Thr Leu Gly Ile Ala Met Ala Ala Phe Leu Val Ser Trp
 245 250 255
 Leu Pro Tyr Leu Val Asp Ala Val Ile Asp Ala Tyr Met Asn Phe
 260 265 270
 Ile Thr Pro Pro Tyr Val Tyr Glu Ile Leu Val Trp Cys Val Tyr
 275 280 285
 Tyr Asn Ser Ala Met Asn Pro Leu Ile Tyr Ala Phe Phe Tyr Gln
 290 295 300
 Trp Phe Gly Lys Ala Ile Lys Leu Ile Val Ser Gly Lys Val Leu
 305 310 315
 Arg Thr Asp Ser Ser Thr Thr Asn Leu Phe Ser Glu Glu Val Glu
 320 325 330
 Thr Asp

<210> 10

<211> 948

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7479890CD1

<400> 10

Met Pro Ser Pro Pro Gly Leu Arg Ala Leu Trp Leu Cys Ala Ala

PI-0236 PCT

1	5	10	15
Leu Cys Ala Ser Arg	Arg Ala Gly Gly Ala	Pro Gln Pro Gly Pro	
20	25	30	
Gly Pro Thr Ala Cys	Pro Ala Pro Cys His	Cys Gln Glu Asp Gly	
35	40	45	
Ile Met Leu Ser Ala	Asp Cys Ser Glu Leu	Gly Leu Ser Ala Val	
50	55	60	
Pro Gly Asp Leu Asp	Pro Leu Thr Ala Tyr	Leu Leu Gly Cys Pro	
65	70	75	
Pro Pro Leu Gln Lys	Ala Gln Ala Val Gly	Gln Leu Gly Glu Tyr	
80	85	90	
Glu Lys Gln Phe Gly	Pro Arg Gln Val Lys	Leu Phe Pro Gln Ser	
95	100	105	
Leu Ser Lys Pro Glu	Leu Ala Cys Glu Val	Pro Ala Asn Leu Pro	
110	115	120	
His Tyr Cys Arg Arg	Leu Asp Ala Asn Leu	Ile Ser Leu Val Pro	
125	130	135	
Glu Arg Ser Phe Glu	Gly Leu Ser Ser Leu	Arg His Leu Trp Leu	
140	145	150	
Asp Asp Asn Ala Leu	Thr Glu Ile Pro Val	Arg Ala Leu Asn Asn	
155	160	165	
Leu Pro Ala Leu Gln	Ala Met Ala Leu Ala	Leu Asn Arg Ile Ser	
170	175	180	
His Ile Pro Asp Tyr	Ala Phe Gln Asn Leu	Thr Ser Leu Val Val	
185	190	195	
Leu His Leu His Asn	Asn Arg Ile Gln His	Leu Gly Thr His Ser	
200	205	210	
Phe Glu Gly Leu His	Asn Leu Glu Thr Leu	Asp Leu Asn Tyr Asn	
215	220	225	
Lys Leu Gln Glu Phe	Pro Val Ala Ile Arg	Thr Leu Gly Arg Leu	
230	235	240	
Gln Glu Leu Gly Phe	His Asn Asn Asn Ile	Lys Ala Ile Pro Glu	
245	250	255	
Lys Ala Phe Met Gly	Asn Pro Leu Leu Gln	Thr Ile His Phe Tyr	
260	265	270	
Asp Asn Thr Ile Gln	Phe Val Gly Arg Ser	Ala Phe Gln Tyr Leu	
275	280	285	
Pro Lys Leu His Thr	Leu Ser Leu Asn Gly	Ala Met Asp Ile Gln	
290	295	300	
Glu Phe Pro Gly Leu	Lys Gly Thr Thr Ser	Leu Glu Ile Leu Thr	
305	310	315	
Leu Thr Arg Ala Gly	Ile Arg Leu Leu Pro	Ser Gly Met Cys Gln	
320	325	330	
Gln Leu Pro Arg Leu	Arg Val Leu Glu Leu	Ser His Asn Gln Ile	
335	340	345	
Glu Glu Leu Pro Ser	Leu His Arg Cys Gln	Lys Leu Glu Glu Ile	
350	355	360	
Gly Leu Gln His Asn	Arg Ile Trp Glu Ile	Gly Ala Asp Thr Phe	
365	370	375	
Ser Gln Leu Ser Ser	Leu Gln Ala Leu Asp	Leu Ser Trp Asn Ala	
380	385	390	
Ile Arg Ser Ile His	Pro Glu Ala Phe Ser	Thr Leu His Ser Leu	

PI-0236 PCT

	395		400		405
Val Lys Leu Asp	Leu Thr Asp Asn Gln	Leu Thr Thr Leu Pro	Leu		
	410		415		420
Ala Gly Leu Gly	Gly Leu Met His Leu	Lys Leu Lys Gly Asn	Leu		
	425		430		435
Ala Leu Ser Gln	Ala Phe Ser Lys Asp	Ser Phe Pro Lys Leu	Arg		
	440		445		450
Ile Leu Glu Val	Pro Tyr Ala Tyr Gln	Cys Cys Pro Tyr Gly	Met		
	455		460		465
Cys Ala Ser Phe	Phe Lys Ala Ser Gly	Gln Trp Glu Ala Glu	Asp		
	470		475		480
Leu His Leu Asp	Asp Glu Glu Ser Ser	Lys Arg Pro Leu Gly	Leu		
	485		490		495
Leu Ala Arg Gln	Ala Glu Asn His Tyr	Asp Gln Asp Leu Asp	Glu		
	500		505		510
Leu Gln Leu Glu	Met Glu Asp Ser Lys	Pro His Pro Ser Val	Gln		
	515		520		525
Cys Ser Pro Thr	Pro Gly Pro Phe Lys	Pro Cys Glu Tyr Leu	Phe		
	530		535		540
Glu Ser Trp Gly	Ile Arg Leu Ala Val	Trp Ala Ile Val Leu	Leu		
	545		550		555
Ser Val Leu Cys	Asn Gly Leu Val Leu	Leu Thr Val Phe Ala	Gly		
	560		565		570
Gly Pro Ala Pro	Leu Pro Pro Val Lys	Phe Val Val Gly Ala	Ile		
	575		580		585
Ala Gly Ala Asn	Thr Leu Thr Gly Ile	Ser Cys Gly Leu Leu	Ala		
	590		595		600
Ser Val Asp Ala	Leu Thr Phe Gly Gln	Phe Ser Glu Tyr Gly	Ala		
	605		610		615
Arg Trp Glu Thr	Gly Leu Gly Cys Arg	Ala Thr Gly Phe Leu	Ala		
	620		625		630
Val Leu Gly Ser	Glu Ala Ser Val Leu	Leu Leu Thr Leu Ala	Ala		
	635		640		645
Val Gln Cys Ser	Val Ser Val Ser Cys	Val Arg Ala Tyr Gly	Lys		
	650		655		660
Ser Pro Ser Leu	Gly Ser Val Arg Ala	Gly Val Leu Gly Cys	Leu		
	665		670		675
Ala Leu Ala Gly	Leu Ala Ala Ala Leu	Pro Leu Ala Ser Val	Gly		
	680		685		690
Glu Tyr Gly Ala	Ser Pro Leu Cys Leu	Pro Tyr Ala Pro Pro	Glu		
	695		700		705
Gly Gln Pro Ala	Ala Leu Gly Phe Thr	Val Ala Leu Val Met	Met		
	710		715		720
Asn Ser Phe Cys	Phe Leu Val Val Ala	Gly Ala Tyr Ile Lys	Leu		
	725		730		735
Tyr Cys Asp Leu	Pro Arg Gly Asp Phe	Glu Ala Val Trp Asp	Cys		
	740		745		750
Ala Met Val Arg	His Val Ala Trp Leu	Ile Phe Ala Asp Gly	Leu		
	755		760		765
Leu Tyr Cys Pro	Val Ala Phe Leu Ser	Phe Ala Ser Met Leu	Gly		
	770		775		780
Leu Phe Pro Val	Thr Pro Glu Ala Val	Lys Ser Val Leu Leu	Val		

PI-0236 PCT

	785		790		795
Val Leu Pro Leu Pro Ala Cys Leu Asn Pro Leu Leu Tyr Leu Leu					
	800		805		810
Phe Asn Pro His Phe Arg Asp Asp Leu Arg Arg Leu Arg Pro Arg					
	815		820		825
Ala Gly Asp Ser Gly Pro Leu Ala Tyr Ala Ala Ala Gly Glu Leu					
	830		835		840
Glu Lys Ser Ser Cys Asp Ser Thr Gln Ala Leu Val Ala Phe Ser					
	845		850		855
Asp Val Asp Leu Ile Leu Glu Ala Ser Glu Ala Gly Arg Pro Pro					
	860		865		870
Gly Leu Glu Thr Tyr Gly Phe Pro Ser Val Thr Leu Ile Ser Cys					
	875		880		885
Gln Gln Pro Gly Ala Pro Arg Leu Glu Gly Ser His Cys Val Glu					
	890		895		900
Pro Glu Gly Asn His Phe Gly Asn Pro Gln Pro Ser Met Asp Gly					
	905		910		915
Glu Leu Leu Leu Arg Ala Glu Gly Ser Thr Pro Ala Gly Gly Gly					
	920		925		930
Leu Ser Gly Gly Gly Gly Phe Gln Pro Ser Gly Leu Ala Phe Ala					
	935		940		945
Ser His Val					

<210> 11

<211> 315

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7482825CD1

<400> 11

Met Glu Ile Val Ser Thr Gly Asn Glu Thr Ile Thr Glu Phe Val			
1	5	10	15
Leu Leu Gly Phe Tyr Asp Ile Pro Glu Leu His Phe Leu Phe Phe			
	20	25	30
Ile Val Phe Thr Ala Val Tyr Val Phe Ile Ile Ile Gly Asn Met			
	35	40	45
Leu Ile Ile Val Ala Val Val Ser Ser Gln Arg Leu His Lys Pro			
	50	55	60
Met Tyr Ile Phe Leu Ala Asn Leu Ser Phe Leu Asp Ile Leu Tyr			
	65	70	75
Thr Ser Ala Val Met Pro Lys Met Leu Glu Gly Phe Leu Gln Glu			
	80	85	90
Ala Thr Ile Ser Val Ala Gly Cys Leu Leu Gln Phe Phe Ile Phe			
	95	100	105
Gly Ser Leu Ala Thr Ala Glu Cys Leu Leu Leu Ala Val Met Ala			
	110	115	120
Tyr Asp Arg Tyr Leu Ala Ile Cys Tyr Pro Leu His Tyr Pro Leu			
	125	130	135

PI-0236 PCT

Leu Met Gly Pro Arg Arg Tyr Met Gly Leu Val Val Thr Thr Trp		
	140	145 150
Leu Ser Gly Phe Val Val Asp Gly Leu Val Val Ala Leu Val Ala		
	155	160 165
Gln Leu Arg Phe Cys Gly Pro Asn His Ile Asp Gln Phe Tyr Cys		
	170	175 180
Asp Phe Met Leu Phe Val Gly Leu Ala Cys Ser Asp Pro Arg Val		
	185	190 195
Ala Gln Val Thr Thr Leu Ile Leu Ser Val Phe Cys Leu Thr Ile		
	200	205 210
Pro Phe Gly Leu Ile Leu Thr Ser Tyr Ala Arg Ile Val Val Ala		
	215	220 225
Val Leu Arg Val Pro Ala Gly Ala Ser Arg Arg Arg Ala Phe Ser		
	230	235 240
Thr Cys Ser Ser His Leu Ala Val Val Thr Thr Phe Tyr Gly Thr		
	245	250 255
Leu Met Ile Phe Tyr Val Ala Pro Ser Ala Val His Ser Gln Leu		
	260	265 270
Leu Ser Lys Val Phe Ser Leu Leu Tyr Thr Val Val Thr Pro Leu		
	275	280 285
Phe Asn Pro Val Ile Tyr Thr Met Arg Asn Lys Glu Val His Gln		
	290	295 300
Ala Leu Arg Lys Ile Leu Cys Ile Lys Gln Thr Glu Thr Leu Asp		
	305	310 315

<210> 12

<211> 312

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7483087CD1

<400> 12

Met Lys Ala Gly Asn Phe Ser Asp Thr Pro Glu Phe Phe Leu Leu		
1	5	10 15
Gly Leu Ser Gly Asp Pro Glu Leu Gln Pro Ile Leu Phe Met Leu		
	20	25 30
Phe Leu Ser Met Tyr Leu Ala Thr Met Leu Gly Asn Leu Leu Ile		
	35	40 45
Ile Leu Ala Val Asn Ser Asp Ser His Leu His Thr Pro Met Tyr		
	50	55 60
Phe Leu Leu Ser Ile Leu Ser Leu Val Asp Ile Cys Phe Thr Ser		
	65	70 75
Thr Thr Met Pro Lys Met Leu Val Asn Ile Gln Ala Gln Ala Gln		
	80	85 90
Ser Ile Asn Tyr Thr Gly Cys Leu Thr Gln Ile Cys Phe Val Leu		
	95	100 105
Val Phe Val Gly Leu Glu Asn Gly Ile Leu Val Met Met Ala Tyr		
	110	115 120

PI-0236 PCT

Asp Arg Phe Val Ala Ile Cys His Pro Leu Arg Tyr Asn Val Ile		
	125	130 135
Met Asn Pro Lys Leu Cys Gly Leu Leu Leu Leu Ser Phe Ile		
	140	145 150
Val Ser Val Leu Asp Ala Leu Leu His Thr Leu Met Val Leu Gln		
	155	160 165
Leu Thr Phe Cys Ile Asp Leu Glu Ile Pro His Phe Phe Cys Glu		
	170	175 180
Leu Ala His Ile Leu Lys Leu Ala Cys Ser Asp Val Leu Ile Asn		
	185	190 195
Asn Ile Leu Val Tyr Leu Val Thr Ser Leu Leu Gly Val Val Pro		
	200	205 210
Leu Ser Gly Ile Ile Phe Ser Tyr Thr Arg Ile Val Ser Ser Val		
	215	220 225
Met Lys Ile Pro Ser Ala Gly Gly Lys Tyr Lys Ala Phe Ser Ile		
	230	235 240
Cys Gly Ser His Leu Ile Val Val Ser Leu Phe Tyr Gly Thr Gly		
	245	250 255
Phe Gly Val Tyr Leu Ser Ser Gly Ala Thr His Ser Ser Arg Lys		
	260	265 270
Gly Ala Ile Ala Ser Val Met Tyr Thr Val Val Thr Pro Met Leu		
	275	280 285
Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Asp Met Leu Lys Ala		
	290	295 300
Leu Arg Lys Leu Ile Ser Arg Ile Pro Ser Phe His		
	305	310

<210> 13

<211> 309

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7483134CD1

<400> 13

Met Gly Ala Lys Asn Asn Val Thr Glu Phe Val Leu Phe Gly Leu		
1	5	10 15
Phe Glu Ser Arg Glu Met Gln His Thr Cys Phe Val Val Phe Phe		
	20	25 30
Leu Phe His Val Leu Thr Val Leu Gly Asn Leu Leu Val Ile Ile		
	35	40 45
Thr Ile Asn Ala Arg Lys Thr Leu Lys Ser Pro Met Tyr Phe Phe		
	50	55 60
Leu Ser Gln Leu Ser Phe Ala Asp Ile Cys Tyr Pro Ser Thr Thr		
	65	70 75
Ile Pro Lys Met Ile Ala Asp Thr Phe Val Glu His Lys Ile Ile		
	80	85 90
Ser Phe Asn Gly Cys Met Thr Gln Leu Phe Ser Ala His Phe Phe		
	95	100 105
Gly Gly Thr Glu Ile Phe Leu Leu Thr Ala Met Ala Tyr Asp Arg		

PI-0236 PCT

	110		115		120
Tyr Val Ala Ile Cys Arg Pro Leu His Tyr Thr Ala Ile Met Asp					
	125		130		135
Cys Arg Lys Cys Gly Leu Leu Ala Gly Ala Ser Trp Leu Ala Gly					
	140		145		150
Phe Leu His Ser Ile Leu Gln Thr Leu Leu Thr Val Gln Leu Pro					
	155		160		165
Phe Cys Gly Pro Asn Glu Ile Asp Asn Phe Phe Cys Asp Val His					
	170		175		180
Pro Leu Leu Lys Leu Ala Cys Ala Asp Thr Tyr Met Val Gly Leu					
	185		190		195
Ile Val Val Ala Asn Ser Gly Met Ile Ser Leu Ala Ser Phe Phe					
	200		205		210
Ile Leu Ile Ile Ser Tyr Val Ile Ile Leu Leu Asn Leu Arg Ser					
	215		220		225
Gln Ser Ser Glu Asp Arg Arg Lys Ala Val Ser Thr Cys Gly Ser					
	230		235		240
His Val Ile Thr Val Leu Leu Val Leu Met Pro Pro Met Phe Met					
	245		250		255
Tyr Ile Arg Pro Ser Thr Thr Leu Ala Ala Asp Lys Leu Ile Ile					
	260		265		270
Leu Phe Asn Ile Val Met Pro Pro Leu Leu Asn Pro Leu Ile Tyr					
	275		280		285
Thr Leu Arg Asn Asn Asp Val Lys Asn Ala Met Arg Lys Leu Phe					
	290		295		300
Arg Val Lys Arg Ser Leu Gly Glu Lys					
	305				

<210> 14

<211> 309

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7478550CD1

<400> 14

Met Met Thr Asn Arg Asn Gln Val Val Leu Gly Arg Met Arg His			
1	5	10	15
Gln Cys Leu Pro Gln Thr Glu Arg Ala His Thr Lys His Asp Leu			
	20	25	30
Ser Leu Gln Ala Gln Leu Gln Gln Lys Val Phe Met Glu Lys Trp			
	35	40	45
Asn His Thr Ser Asn Asp Phe Ile Leu Leu Gly Leu Leu Pro Pro			
	50	55	60
Asn Gln Thr Gly Ile Phe Leu Leu Cys Leu Ile Ile Leu Ile Phe			
	65	70	75
Phe Leu Ala Ser Val Gly Asn Ser Ala Met Ile His Leu Ile His			
	80	85	90
Val Asp Pro Arg Leu His Thr Pro Met Tyr Phe Leu Leu Ser Gln			
	95	100	105

PI-0236 PCT

```

Leu Ser Leu Met Asp Leu Met Tyr Ile Ser Thr Thr Val Pro Lys
    110                      115                      120
Met Ala Tyr Asn Phe Leu Ser Gly Gln Lys Gly Ile Ser Phe Leu
    125                      130                      135
Gly Cys Gly Val Gln Ser Phe Phe Phe Leu Thr Met Ala Cys Ser
    140                      145                      150
Glu Gly Leu Leu Leu Thr Ser Met Ala Tyr Asp Arg Tyr Leu Ala
    155                      160                      165
Ile Cys His Ser Leu Tyr Tyr Pro Ile Arg Met Ser Lys Met Met
    170                      175                      180
Cys Val Lys Met Ile Gly Gly Ser Trp Thr Leu Gly Ser Ile Asn
    185                      190                      195
Ser Leu Ala His Thr Val Phe Ala Leu His Ile Pro Tyr Cys Arg
    200                      205                      210
Ser Arg Ala Ile Asp His Phe Phe Cys Asp Val Pro Ala Met Leu
    215                      220                      225
Leu Leu Ala Cys Thr Asp Thr Trp Val Tyr Glu Tyr Met Val Phe
    230                      235                      240
Val Ser Thr Ser Leu Phe Leu Leu Phe Pro Phe Ile Gly Ile Thr
    245                      250                      255
Ser Ser Cys Gly Arg Val Leu Phe Ala Val Tyr His Met His Ser
    260                      265                      270
Lys Glu Gly Arg Lys Lys Ala Phe Thr Thr Ile Ser Thr His Leu
    275                      280                      285
Thr Val Val Ile Phe Tyr Tyr Ala Pro Phe Val Tyr Thr Tyr Leu
    290                      295                      300
Arg Pro Thr Glu Ser Pro Leu Thr Ser
    305

```

<210> 15

<211> 315

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7483142CD1

<400> 15

```

Met Ser Ile Thr Lys Ala Trp Asn Ser Ser Ser Val Thr Met Phe
  1                      5                      10                      15
Ile Leu Leu Gly Phe Thr Asp His Pro Glu Leu Gln Ala Leu Leu
    20                      25                      30
Phe Val Thr Phe Leu Gly Ile Tyr Leu Thr Thr Leu Ala Trp Asn
    35                      40                      45
Leu Ala Leu Ile Phe Leu Ile Arg Gly Asp Thr His Leu His Thr
    50                      55                      60
Pro Met Tyr Phe Phe Leu Ser Asn Leu Ser Phe Ile Asp Ile Cys
    65                      70                      75
Tyr Ser Ser Ala Val Ala Pro Asn Met Leu Thr Asp Phe Phe Trp
    80                      85                      90
Glu Gln Lys Thr Ile Ser Phe Val Gly Cys Ala Ala Gln Phe Phe

```

PI-0236 PCT

	95		100		105
Phe Phe Val Gly Met Gly Leu Ser Glu Cys Leu Leu Leu Thr Ala					
	110		115		120
Met Ala Tyr Asp Arg Tyr Ala Ala Ile Ser Ser Pro Leu Leu Tyr					
	125		130		135
Pro Thr Ile Met Thr Gln Gly Leu Cys Thr Arg Met Val Val Gly					
	140		145		150
Ala Tyr Val Gly Gly Phe Leu Ser Ser Leu Ile Gln Ala Ser Ser					
	155		160		165
Ile Phe Arg Leu His Phe Cys Gly Pro Asn Ile Ile Asn His Phe					
	170		175		180
Phe Cys Asp Leu Pro Pro Val Leu Ala Leu Ser Cys Ser Asp Thr					
	185		190		195
Phe Leu Ser Gln Val Val Asn Phe Leu Val Val Val Thr Val Gly					
	200		205		210
Gly Thr Ser Phe Leu Gln Leu Leu Ile Ser Tyr Gly Tyr Ile Val					
	215		220		225
Ser Ala Val Leu Lys Ile Pro Ser Ala Glu Gly Arg Trp Lys Ala					
	230		235		240
Cys Asn Thr Cys Ala Ser His Leu Met Val Val Thr Leu Leu Phe					
	245		250		255
Gly Thr Ala Leu Phe Val Tyr Leu Arg Pro Ser Ser Ser Tyr Leu					
	260		265		270
Leu Gly Arg Asp Lys Val Val Ser Val Phe Tyr Ser Leu Val Ile					
	275		280		285
Pro Met Leu Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Glu Ile					
	290		295		300
Lys Asp Ala Leu Trp Lys Val Leu Glu Arg Lys Lys Val Phe Ser					
	305		310		315

<210> 16

<211> 307

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7483151CD1

<400> 16

Met Ala Glu Glu Asn Lys Ile Leu Val Thr His Phe Val Leu Thr					
1	5		10		15
Gly Leu Thr Asp His Pro Gly Leu Gln Ala Pro Leu Phe Leu Val					
	20		25		30
Phe Leu Val Ile Tyr Leu Ile Thr Leu Val Gly Asn Leu Gly Leu					
	35		40		45
Met Ala Leu Ile Trp Lys Asp Pro His Leu His Thr Pro Ile Tyr					
	50		55		60
Leu Phe Leu Gly Ser Leu Ala Phe Ala Asp Ala Cys Thr Ser Ser					
	65		70		75
Ser Val Thr Ser Lys Met Leu Ser Ile Phe Leu Ser Lys Asn His					

PI-0236 PCT

	80		85		90									
Met	Leu	Ser	Met	Ala	Lys	Cys	Ala	Thr	Gln	Phe	Tyr	Phe	Phe	Gly
	95		100		105									
Ser	Asn	Ala	Thr	Thr	Glu	Cys	Phe	Leu	Leu	Val	Val	Met	Ala	Tyr
	110		115		120									
Asp	Arg	Tyr	Val	Ala	Ile	Cys	Asn	Pro	Leu	Leu	Tyr	Pro	Val	Val
	125		130		135									
Met	Ser	Asn	Ser	Leu	Cys	Thr	Gln	Phe	Ile	Gly	Ile	Ser	Tyr	Phe
	140		145		150									
Ile	Gly	Phe	Leu	His	Ser	Ala	Ile	His	Val	Gly	Leu	Leu	Phe	Arg
	155		160		165									
Leu	Thr	Phe	Cys	Arg	Ser	Asn	Ile	Ile	His	Tyr	Phe	Tyr	Cys	Glu
	170		175		180									
Ile	Leu	Gln	Leu	Phe	Lys	Ile	Ser	Cys	Thr	Asn	Pro	Thr	Val	Asn
	185		190		195									
Ile	Leu	Leu	Ile	Phe	Ile	Phe	Ser	Ala	Phe	Ile	Gln	Val	Phe	Thr
	200		205		210									
Phe	Met	Thr	Leu	Ile	Val	Ser	Tyr	Ser	Tyr	Ile	Leu	Ser	Ala	Ile
	215		220		225									
Leu	Lys	Lys	Lys	Ser	Glu	Lys	Gly	Arg	Ser	Lys	Ala	Phe	Ser	Thr
	230		235		240									
Cys	Ser	Ala	His	Leu	Leu	Ser	Val	Ser	Leu	Phe	Tyr	Gly	Thr	Leu
	245		250		255									
Phe	Phe	Met	Tyr	Val	Ser	Ser	Arg	Ser	Gly	Ser	Ala	Ala	Asp	Gln
	260		265		270									
Ala	Lys	Met	Tyr	Ser	Leu	Phe	Tyr	Thr	Ile	Ile	Ile	Pro	Leu	Leu
	275		280		285									
Asn	Pro	Phe	Ile	Tyr	Ser	Leu	Arg	Asn	Lys	Glu	Val	Ile	Asp	Ala
	290		295		300									
Leu	Arg	Arg	Ile	Met	Lys	Lys								
	305													

<210> 17

<211> 2422

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2536292CB1

<400> 17

```

gacaagggtc tcactttggt gccaggttg gtctcgaaact ccaggtctcg agcggtcctc 60
ctgttttggc ctcccaaagc attgggatta caggcatgca ccacaattcc tgacctacag 120
ttttcttctt atgggctttg tacatttctt gcctcctggt atttttaagg aaatgtaaga 180
gtgtgaatca ggtactgatg gttttgactt cattatgatc agcatctttg gcttccatca 240
gagtagacac tgaacgtttt catccaaccg gatgatgagt gcattagaac tacataagag 300
atgaaatcct ttctccctgg gacatgtatc cttttgtggt ctgcttttaa tttgatgttt 360
ttcagtcttt tcaggctgaa atataatatt tgtatcattc tgagagcctg taacaccatg 420
ctttccagca atacaattat ggaaattttt ttcctctctc atattgacat tgggatttgg 480
aggaacttac tgctactcct gatgcctatt tacacctttt tgatatgtcc ccagcagaag 540
aagcccatgg gcttgctttt ccttcacttg tctgttgcta atacgatgac acttctccgc 600

```

PI-0236 PCT

```

aaagttatcc cattggcagt aaaatctttc aacactaaaa atcttttgaa ttatactgga 660
tgtaggggaat ttgaattttt atatagagta tcttgggggac ttcccttatg tacgacatac 720
ctcctaagca tgggtgcaggc cctccgtggg agccccagca aatccagggt catcaacagt 780
ctcatctaca tcaagcttgt gccatttgtt gacactatca aatatggcag tgtcaccaag 840
aatctctcca taaaaatgtg tttagctaca ccacacatgg gcaacacaat tgcgtgtctc 900
catacgagtg ttatcacatt ccaggactta atttttctgg ttctcatgag ttgagccagt 960
ggctacttgg tgattttcct tcacagacac cagaaaaata atccacatct tcacaacaac 1020
agctgttctc ctatagcctc ccatgagacc ggaaccatca cgactgcgct gctgcttatg 1080
atttgcttcg ttgtatttaa tgtgagcaac tcgtgccaca gcatttacct aagtacggtg 1140
aagaaaaggg atcagttgtg gaccatctca gatttgattt cctcatgtta ccccatTTTT 1200
tgggccattt ttgctcattg gtagagaaag tcttatcctg aattcaaaat ctatgaagta 1260
gagaaagtcg tcttatccca tagatatgtc aaagtaaact ttcaccaata taaactTTTT 1320
ccaaaacaac attctataaa gagttggtat tctgaagaaa tgatgggatt taggtctgaa 1380
aaatgagata tttctcagtt cactgcatga aatataatct aatgacctt accttcagta 1440
aagacaatat tgcatagctt agtatttaat gtttgtatac atagatataa tcatttaata 1500
gacaagtaga tagataattt ttcatgtcag atgtttataat ggacttcttg cactagtcac 1560
tcaatatcca tatttttctt actttaattt gtgttcagca tgaaaaatca tttcaaaagg 1620
agatcagggg ctccaggaag cagagcgtaa ctccccactt ctaaactgtg ggcttcacac 1680
aatgacatcc taccaagaaa tcagtatagc aaaggaggaa aaagagtaac ttcgcatttg 1740
agaaacctaa caaacactgt ctccagccagg tgatcaaggc cactgtgagc agtgatcagt 1800
caagtcgaaa gtgcataatc tatgacagga tgagaatggc attttacctc gacatcttcc 1860
ctcccaaac taataatcct actacatcag acaaacctca gctgagaaac agtttccaaa 1920
aacacctgac tagtaccctc taaaaccatc aagggtcatca agaacaatgt agtcctgaga 1980
aatgatcaca gccgggaaga acctaaggac ttggtgtcct ttggtatcct gatatggaat 2040
cctggagcag aaaaggacat taagtaaaaa ccaaggaaat agaaataaag tataaactta 2100
ataatatatc aatatattt catttaattgt gacaaatgta acatagtaat gtaagacatt 2160
aaaatgggga aaattgggta tgatgtggag gagaactctg tgtattctat ttgcaactct 2220
tctgtaaatg taaacctatt tgaacaatta aaaattttat tttttcaaaa caaaaaaaa 2280
aaaacaaaac aacaacaaca caaaaacaca gaggggcgcg gcgaccaaa atatcaaacg 2340
gccaccgcg gggggggcgc cccaccata aagagataaa cacacacgga ggtaaaaaac 2400
gggggaaagc ggtccctctc gc 2422

```

<210> 18

<211> 1912

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7477708CB1

<400> 18

```

aggaaacaag aaagtcagct gtgtggaaga caatgactca tatacttttg ctgtactact 60
tggtgtttct tttgccaca gagtcctgta ggacattgta tcaggtgtat gcatggtgt 120
ctgtacagac tactcccagt gtactcaacc ttgccctcca gacactcagg gaaatatggg 180
gttttcatgc aggcaaaaga catggcacia gatcactgac acctgccgga ctcttaatgc 240
cctcaacatc tttgaggagg attcacgttt ggttcagcca tttgaagaca atataaaaat 300
aagtgtatat actggaaagt ctgagaccat aacagatatg ttgctacaaa agtgtccac 360
agatctgtct tgtgtaatta gaaacattca gcagtctccc tggataccag gaaacattgc 420
cgtaattgtg cagctcttac acaacatc aacagcaata tggacagggt ttgatagggc 480
aaagatgcag agttacagca ccatagccaa ccacattctt aacagcaaaa gcatctccaa 540
ctggactttc attcctgaca gaaacagcag ctatatcctg ctacattcag tcaactcctt 600

```


PI-0236 PCT

tgcaagaagg ctattcatag ataacatccc tgttgacata tcagatgtct tcattcatac 660
tatgggcacc accatatctg gagataacat tggaaaaaat ttcacttttt ctatgagaat 720
taatgacacc agcaatgaag tcaactgggag agtgttgatc agcagagatg aacttcggaa 780
ggtgccttcc ccttctcagg tcatcagcat tgcatttcca actattgggg ctattttgga 840
agccagtctt ttggaaaatg ttactgtaaa tgggcttgtc ctgtctgcc a ttttgcccaa 900
ggaacttaaa agaatctcac tgatttttga aaagatcagc aagtcagagg agaggaggac 960
acagtgtgtt ggctggcact ctgtggagaa cagatgggac cagcaggcct gcaaaatgat 1020
tcaagaaaac tcccagcaag ctgtttgcaa atgtaggcca agcaaattgt ttacctcttt 1080
ctcaattctt atgtcacctc acatcttaga gagtctgatt ctgacttaca tcacatatgt 1140
aggcctgggc atttctattt gcagcctgat cctttgcttg tccattgagg tcctagtctg 1200
gagccaagtg acaaagacag agatcaccta tttacgccat gtgtgcattg ttaacattgc 1260
agccactttg ctgatggcag atgtgtggtt cattgtggtt tcctttctta gtggcccaat 1320
aacacaccac aagggatgtg tggcagccac attttttgtt catttctttt acctttctgt 1380
atttttctgg atgcttgcca aggcactcct tctcctctat ggaatcatga ttgttttcca 1440
taccttgccc aagtcagtc tgggtggcatc tctgttttca gtgggctatg gatgcccttt 1500
ggccattgct gccatcactg ttgctgccac tgaacctggc aaaggctatc tacgacctga 1560
gatctgctgg ctcaactggg acatgaccaa agccctcctg gccttcgtga tcccagcttt 1620
ggccatcgtg gtagtaaacc tgatcacagt cacactggtg attgtcaaga cccagcgagc 1680
tgccattggc aattccatgt tccaggaagt gagagccatt gtgagaatca gcaagaacat 1740
cgccatcctc acaccacttc tgggactgac ctggggattt ggagtagcca ctgtcatcga 1800
tgacagatcc ctggccttcc acattatctt ctcctgtctc aatgcattcc aggtaagtcc 1860
agatgcttct gaccaagtgc aaagtgagag aattcatgaa gatgttctgt ga 1912

<210> 19

<211> 1326

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7474823CB1

<400> 19

atggttctag gtaaaaatgt ctccatgtct ggcccaaggc cagcatcatg gcagtccac 60
cctcaagggc ttgagcttgt ctttgaaaaa tggccctgca gatgtagtta ttcagcagtc 120
ctggtgatct cctccatcag cagttctgga gctggggaca tcccagatca agattctggc 180
caatattggt tctgatgag ggctgtcttt ctggcttgca gacggctgcc gtctacctgc 240
gtcctcaaga ggcctttctc tgagtgcgca cagagagaga gaactaattt ggttctcatg 300
aaaaaatggg aattcctgga agtacctgat acatttgaag taactcaaca aagtgtgac 360
tccattcctt tgtacatccc tcacacgctg ttcgaatggg attttgga a ggaaatctgt 420
gtattttggc tcaactactga ctatctgtta tgtacagcat ctgtatataa cattgtcctc 480
atcagctatg atcgatacct gtcagtctca aatgctgtgt cttatagaac tcaacatact 540
ggggtcttga agattgttac tctgatgggt gccgtttggg tgctggcctt cttagtgaat 600
gggccaatga ttctagtttc agagtcttgg aaggatgaag gtagtgaatg tgaacctgga 660
tttttttcgg aatggtacat ccttgccatc acatcattct tggaaatcgt gatcccagtc 720
atcttagtcg cttatttcaa catgaatatt tattggagcc tgtggaagcg tgatcatctc 780
aggcttgggc atcccaaggg atggggccag ctggtgtcca gactgccaca tggggttgag 840
ggacagccgt ggccgctgca gctggtccct cggatggggg acattgaagt aggaggcttg 900

ERROR: ioerror
OFFENDING COMMAND: image

STACK:

false
-savelevel-